Universidade de Aveiro Departamento de Electrónica, Telecomunicações e Informática
2014

Duarte Santos

Controlo de Congestionamento em Redes Inter-tecnologia

Universidade de Aveiro
2014

Departamento de
Electrónica, Telecomunicações e Informática

Duarte
Santos

Controlo de Congestionamento em Redes Inter-tecnologia

"Intellectual growth should commence at birth and cease only at death."

— Albert Einstein

Universidade de Aveiro 2014

Departamento de Electrónica, Telecomunicações e Informática

Duarte Santos

Controlo de Congestionamento em Redes Inter-tecnologia

Dissertação apresentada à Universidade de Aveiro para cumprimento dos requesitos necessários à obtenção do grau de Mestre em Engenharia Electrónica e Telecomunicações, realizada sob a orientação científica da Doutora Susana Sargento, Professora auxiliar do Departamento de Electrónica Telecomunicações e Informática da Universidade de Aveiro e co-orientação do Doutor Pedro Neves, PT Inovação.

o júri / the jury

presidente / president	**Professor Doutor José Rodrigues Ferreira da Rocha** Professor Catedrático do Departamento de Eletrónica, Telecomunicações e Informática da Universidade de Aveiro
vogais / examiners committee	**Professora Doutora Marília Pascoal Curado** Professora Auxiliar do Departamento de Informática da Universidade de Coimbra (Arguente)
	Professora Doutora Susana Isabel Barreto de Miranda Sargento Professora Auxiliar do Departamento de Eletrónica, Telecomunicações e Informática da Universidade de Aveiro (orientadora)

agradecimentos / acknowledgements

Gostaria de agradecer aos meus pais e irmão por me terem proporcionado a possibilidade de tirar um curso superior e por todo o apoio prestado ao longo destes anos. Aproveito também para agradecer à minha namorada, Filipa, pela paciência e o carinho dados durante este percurso.

Agradeço também a todos os meus amigos que sempre me apoiaram, em especial ao grupo de redes que partilhou comigo a etapa final desta caminhada.

Deixo também um grande agradecimento para os Engenheiros João Nogueira e Tiago Cardoso pela ajuda e tempo disponibilizado e um agradecimento geral para a PT Inovação pelo apoio prestado.

Em último, mas não menos importante, agradeço à Professora Susana Sargento pela oportunidade que me proporcionou de desenvolver a tese de mestrado na área das redes de telecomunicações. Muito obrigado.

Resumo

A necessidade de estar sempre contatável e informado do que se passa à sua volta tornou-se uma realidade na vida da maioria das pessoas, o que é uma causa da grande necessidade de maior capacidade por parte das redes. A natural globalização da sociedade acompanhada pelo uso intensivo de redes sociais, em que partilhar e visualizar conteúdos é uma constante, torna necessária uma adaptação por parte das redes móveis.

Este aumentar da necessidade por parte dos utilizadores de uma maior capacidade das redes móveis levanta problemas aos operadores. Muitas vezes este aumento da procura não é acompanhado pela evolução das tecnologias disponíveis. Esta situação traduz-se em problemas nas redes de acesso que causam prejuízo para o operador e falhas no serviço prestado ao cliente. É de todo importante que o serviço seja o melhor possível tanto para o cliente como para o operador, pelo que impedir e solucionar o congestionamento na rede é do interesse de todos.

A qualidade do serviço prestado é também um ponto importante e de grande preocupação. Garantir que os serviços oferecidos são recebidos com as melhores condições é uma preocupação por parte dos prestadores de serviços. Para tal é necessário conseguir avaliar o serviço fornecido de forma determinística, a fim de conseguir a melhor satisfação por parte do utilizador.

Contudo, muitos desafios afectam ainda as redes móveis atuais. Esta Dissertação propõe-se a identificar e solucionar casos de congestionamento nas redes móveis, bem como aferir qual a qualidade dos serviços prestados ao cliente. Para tal, irá ser criado um módulo, Metrics Evaluation, capaz de ser integrado na arquitectura actual da rede que irá detectar casos de congestionamento nas redes de acesso, para que estes possam ser resolvidos impondo regras na rede usando o PCRF. Serão também avaliados quais os melhores métodos para aferir a qualidade de serviço prestado ao cliente, recorrendo ao desenvolvimento de uma probe que irá inferir o estado do serviço Meo Go numa determinada célula e comunicará os seus resultados ao PCRF para que, no caso em que é experienciada uma baixa qualidade, este possa interferir e melhorar a QoS.

Serão então desenvolvidos cenários que demonstram a utilidade dos mecanismos de deteção, resolução do congestionamento e obtenção da qualidade de serviço do utilizador. Após testados e validados em diversos cenários, os mecanismos desenvolvidos serão usados em redes reais no contexto do operador, para oferecer uma maior disponibildade da rede para as exigências dos utilizadores bem como proporcionar um melhor serviço da mesma.

Abstract

The need to be always contactable and informed of what is happening around has become a reality in most people's lives, which is also a cause for the need for a greater capacity of the networks. The natural globalization of society accompanied by intensive use of social networks, where share and content visualization is a constant need, requires an adjustment by the mobile networks.

This increase of the need by the users for a greater capacity of mobile networks raises problems for operators. This increase of demand for network capacity is not accompanied by the evolution of the technologies available. This represents problems in access networks that cause losses to the operator and failures in the customer service. It is quite important that the service is the best possible for both the client and the operator; prevent and resolve network congestion is of everyone's interest.

The quality of service is also an important point of great concern. Ensure that the services offered are received in the best conditions is a concern of service providers. To achieve this, it is necessary to evaluate the service provided in a deterministic way, in order to achieve the best satisfaction from the user.

However, many challenges still need to be surpassed on current mobile networks. This Dissertation proposes to identify and resolve cases of congestion in mobile networks, as well as infer the quality of services provided to the customer. To do so, it will developed a module, Metrics Evaluation, capable of being integrated in the actual network that will detect cases of congestion in radio access networks, so that it can be resolved forcing rules in the network using PCRF. It will also be evaluated the best methods for achieving the quality of customer service by developing a probe that will give the state of the service Meo Go in a determined cell and communicate the results to PCRF, so in the case of bad quality is detected, it can act and improve QoS .

It will then be developed scenarios that demonstrate the usefulness of the detection, resolution of congestion and achieving the quality of service of the user. Once tested and validated in various scenarios, our developed approach will be later used in real networks in the context of the operator, to offer a greater availability of the network to the requirements of users and provide a better service than before.

Contents

List of Figures

List of Tables

Acronyms

3GPP	3rd Generation Partnership Project
AF	Application Function
ALCAP	Access Link Control Application Protocol
AMR	Adaptive Multi-Rate
ARIB	Association of Radio Industries and Businesses
ATIS	Alliance for Telecommunications Industry Solutions
BTS	Base Transceiver Station
CAC	Call Admission Control
CCSA	China Communications Standards Association
CDR	Charging Data Record
CE	Channel Element
CIR	Carrier-to-interference ratio
CS	Circuit Switched
DPI	Deep Packet Inspection
EMS	Element Management Systems
EPC	Evolved Packet Core
ETSI	European Telecommunications Standards Institute
GPRS	General Packet Radio Service
GSM	Global System for Mobile Communications
HLR	Home Location Register
HSPA	High Speed Packet Access

HSDPA	High Speed Downlink Packet Access
HSS	Home Subscriber Server
HSUPA	High Speed Uplink Packet Access
HTTP	Hypertext Transfer Protocol
IP	Internet Protocol
IPM	IpRaft Policy Manager
ITU	International Telecommunications Union
KPI	Key Performance Indicator
KQI	Key Quality Indicator
LTE	Long Term Evolution
QoE	Quality of Experience
QoS	Quality of Service
NBAP	NodeB Application Part
OCS	Online Charging System
OFCS	Off-line Charging System
OF	Mobility Management Entity
PCRF	Policy and Charging Rules Function
PCC	Policy and Charging Control
PS	Packet Switch
QAM	Quadrature Amplitude Modulation
QPSK	Quadrature Phase Shift Keying
QRE	Quantum Rule Engine
RAB	Radio Access Bearer
RAN	Radio Access Network
RANAP	Radio Access Network Application Part
RAT	Radio Access Technology
RF	Radio Frequency

RNC	Radio Network Controller
RNSAP	Radio Network Subsystem Application Part
RTT	Round Trip Time
SLA	Service Level Agreement
SGSN	Serving GPRS Support Node
SPR	Subscriber Profile Repository
TDF	Traffic Detection Function
TTA	Telecommunications Technology Association
TTC	Telecommunication Technology Committee
TTI	Transmission Time Interval
UE	User Equipment
UMTS	Universal Mobile Telecommunications System
USB	Universal Serial Bus
UTRAN	Universal Terrestrial Radio Access Network
VOD	Video on Demand
W-CDMA	Wideband Code Division Multiple Access
XML	eXtensible Markup Language

Chapter 1

Introduction

1.1 Motivation

Nowadays, with the development of mobile devices, that have functions beyond voice and text communications, providing services like multimedia content, video and online gaming which grow in popularity, it is caused an exponential growth of traffic in mobile networks. Although the data capacity of networks has increased, the growth of user traffic outpace the growth in capacity. This exponential growth of cellular data is a big concern to the mobile network operators'; therefore, the efficient use of all available resources is extremely important. The actions of detecting, avoiding and controlling congestion scenarios with the best solutions are important to maximize resources and provide the best service to the users.

To detect and prevent congestion, it is required to take into account two different stages. First of all, it is necessary to detect that there are problems in the network, specifically in the radio access network (where congestion occurs). After the detection, it is necessary an evaluation of which actions can be taken. This starts by knowing who are the main contributors to the problem and decide how they will be affected. There are different clients in the network, and network providers may have different policies to choose the type of clients to be affected when congestion occurs.

The first stage must evaluate the network status at its lowest level, directly in the NodeBs and in Radio Network Controller (RNCs). At this level, many factors can influence the status of the network. In 3G for example, factors like the lack code availability or the lack of power resources can cause congestion. These factors must be chosen and evaluated to deduce a qualitative value for the state of network.

Network providers are also interested in knowing the quality of the services that they provide. Detecting that their service is being received with bad quality is a great concern for them. Therefore, it will be created a probe that is capable of evaluating the status of the service Meo Go in the cell; and in the case of low QoS, it will report the problem to the Policy Charging Rules Function (PCRF), so the PCRF can act upon the congestion. This type of features allow operators to create new business opportunities like extra charging for QoS guarantees in specific services.

The second stage consists in identifying the main contributors of the congestion, the called heavy users, and act directly on the service of these users to relieve the load on the congested cell. The actions to take upon these users must take into account the type of subscribed services by the clients. Clients with a better subscribed service must have a better service than clients with a lower service plan, according to the operators policies.

After all the information about the cell status is evaluated, the result of that evaluation must be passed to the network element that will force the changes into the costumers services. To achieve this, is necessary to create an interface with the existent platform, capable of receiving and analysing the information concerning the network. After that step, it is possible to create new policies in the network in the PCRF module. PCRF is the module in charge of the creation of dynamical rules that can be forced in the network to stop and avoid congestion.

The last step in this operation is the evaluation of the received data. Rules must be constructed, taking into the account the information received in the report describing the congestion level of the network in a specific cell. According to the severity of the information received the rules must adapt the severity of its actions. The new policies that result from the rules will affect users in the congested cell. The downgrade of the type of service or even drop of low priority clients are some of the actions considered in the rulesets to stop congestion.

All the modules described will be constructed or enhanced to support the new features described before. The modules will then be tested and the results of their operation will be described in this Dissertation.

1.2 Objectives

Taking into account the fragilities of the radio access network exposed before, the main objective for this Dissertation is to gather congestion metrics of the network, evaluate the QoS provided to the services, and decide on how to react to congestion situations. To achieve this main goal, the objectives that must be fulfilled to accomplish are the following:

- Evaluate the network status: in order to discover the parameters that determine if the network is congested;
- Determine the level of congestion: different actions must be taken according to the level of congestion so it can be solved; therefore it is necessary to determine the level of congestion;
- Evaluate the QoS of the services: in order to give the best quality of the service, it is crucial to be able to determine the QoS of the services;
- Communicate the result of the evaluation: it is necessary to send the result of the network evaluation to the module that is capable of imposing actions on the network that can solve the congestion;

- Choose the actions to be taken: To resolve congestion it is necessary that a set of actions be chosen;
- Implement the actions: To achieve the result of the actions, they must be implemented by in the network a specific module;
- Evaluate the Solution: Results must be taken to understand the effects of the actions.

1.3 Organization

This Dissertation is divided in seven chapters organized as follows:

- Chapter 1: presents the motivation of the Dissertation and contextualizes it in the environment where it will be developed.
- Chapter 2: presents the state of the art of the mobile networks, describing the architecture of the network and the wireless technologies.
- Chapter 3: shows the existing architecture and how the solution will work and is integrated on the architecture.
- Chapter 4: describes the software implementation required to evaluate the network and to give response to its problems.
- Chapter 5: describes the construction of the Meo Go probe and all its requirements.
- Chapter 6: shows the testing scenarios of the solution and its results.
- Chapter 7: summarizes the work done on this Dissertation and the possible future improvements.

Chapter 2

State of the Art

2.1 Introduction

To understand the solution that is being proposed it is important to know the architecture of its elements and the technologies that are being used. In this Dissertation it is proposed a solution that extends network elements. Of the 3GPP architecture, so this state of the art starts by contextualize the network architecture and describe its features. After that, it is described the radio access network network architecture and the access technologies. Since congestion is the main concern of this Dissertation, it is described and shown the main contributors to congest the network. After that, RAN evaluation metrics are described in detail, since they are one of the ways to evaluate the state of the network. Probing and DPI are also described in this section since they are methods to achieve rich information about the network. This chapter is organized as follows:

- Section 2.2 describes the cellular technologies and its upgrades and evolutions.
- Section 2.3 describes the architecture of the 3GPP network, including Policy and Charging Control (PCC), radio access network architecture and radio access technologies.
- Section 2.4 shows the definition of congestion in cellular networks and the main elements that can cause congestion.
- Section 2.5 describes the network metrics that evaluate determined parameters of the radio access networks. It is also shown their usefulness and how they are going to be used in this Dissertation.
- Section 2.6 shows the concept of probing, including the concept of DPI and its utilization.
- Section 2.7 describes the functioning of Call Admission Control solution and solutions proposed and implemented in the literature.
- Section 2.8 depicts the conclusions and the summary of the full chapter.

2.2 Access Technologies

In this section it will be described the radio access technologies that are used in this Dissertation. The radio access technologies have been extend over the years having increased its capacities and features, mainly developing the downlink and uplink bandwidth. This section describes the 3G and 4G technologies with the various developments of both technologies.

2.2.1 3G

3G is the acronym for third generation of wireless technology used in mobile telecommunication networks. In 1988 ITU-R defined the requirements for the 3G technology. It is the evolution and successor of the 2G technology and it has several enhancements over the previous technology (2G), like high speed transmission, advanced multimedia access and global roaming. In 2001 the first commercial service of 3G started; its networks enable the operators to give a wider range of advanced services and improve the network capacity due to the increase of spectral efficiency. Some of the 3G requirements are the following:

- To integrate fixed and mobile telephony, Internet and broadcast TV;
- To provide access to anything, anywhere and anytime;
- To support audio, video, data, 'multimedia';
- To support higher data rates than 2G:
  - More than 2Mbit/s for indoor, low-range outdoor with maximum speed of 10km/h;
  - More than 384kbit/s for suburban outdoor with maximum speed of 120km/h;
  - More than 144kbit/s for rural outdoor with maximum speed of 500km/h;
- Easy handover between 3G and 2G;

2.2.1.1 Universal Mobile Telecommunications System

Universal Mobile Telecommunications System (UMTS) is one of the 3G technologies for mobile networks and it is an evolution of the GSM/GPRS. Both UMTS and GSM have similar elements, which permitted the transition from GSM to UMTS to be easier and cheaper. UMTS was developed and financed by 3GPP and it specifies the radio access network (UTRAN) and also the core network. The modulation scheme used by UMTS is Wideband Code-Division Multiple Access (W-CDMA), it can use a pair or unpaired 5MHz wide channels with a central frequency around 2GHz. To achieve spectral efficiency and coexist with others radio access techniques UMTS uses Frequency Division Duplex (FDD) or Time Division Duplex (TDD) variants permitting the use of a wide range of bands.

2.2.1.2 High Speed Packet Access

High Speed Packet Access is an evolution to the standard 3G that technology that comprehends two telecommunications protocols HSDPA for downlink connections (3GPP release 5) and HSUPA for uplink connection (3GPP release 6). The scope of this evolution is to enhance the performance of the base 3G telecommunications technology, mainly due to the increase of packet data performance of both uplink and downlink connections. The HSPA was higlhy adopted by mobile telecommunications operators due to the low cost of actualization from the existing 3G network, and the efficient use of spectrum that makes its cost per megabyte of data delivered cheaper than before. These factors made it the most widely deployed mobile technology around the world.

2.2.1.3 High Speed Downlink Packet Access

The High Speed Downlink Packet Access is a protocol for mobile telecommunications data transmission that enhances the downlink connections of the standard UMTS technology. It is known as 3.5G due to be an evolution for 3G networks. HSDPA is in charge of upgrading the downlink connection, it doubles the capacity of the network and increases download data speeds reaching theoretically values of 14 Mbps. This enhancement is very useful for applications that demand for heavy data transmissions such as streaming or mobile TV. The natural need for higher data speeds on the downlink (due to the unbalanced traffic of users) made 3GPP develop and release the HSDPA on release 5; after that it was developed the uplink connection HSUPA that was released later.

The improvement of the data rate transmission in the downlink connection is due to the use of different techniques for modulation and coding. HSDPA uses different channels than W-CDMA, it uses a new physical channel (HS-PDSCH) and a new transport channel (HS-DSCH). This channels allow higher data speeds compared to te conventional UMTS, due to allowing two modes of coding (QPSK and 16-QAM), turbo codification, a shorter transmission time interval (TTI) and a fast link adaptation to the channel conditions.

2.2.1.4 High Speed Uplink Packet Access

The High Speed Uplink Packet Access, contrasting with HSDPA, is a mobile telecommunications protocol that enhances the capabilities of UMTS networks for uplink connections (from UE to NodeB). This is the second most important updated to W-CDMA and it was introduced by 3GPP in release 6. HSUPA allows a considerable increase of data speed transfer to the uplink connection, compared to the standard UMTS, with theoretical values of 5.8Mbps. Despite of being a great advantage to the uplink connection, it is considerable smaller than the data speeds achieved in the downlink connections.

The increase in the uplink direction is mainly due to the addition of a new transport channel called Enhanced Dedicated Channel (E-DCH). Some of the principal characteristics introduced by this new channel are the fast Node B scheduling that reduces the latency, the use of fast HARQ (Hybrid Automatic Repeat Request) and a shorter transmission time interval. These factors make the uplink connection faster, allowing the use of applications that require more throughput of the uplink, like sending large emails, streaming or use VoIP.

2.2.1.5 High Speed Packet Access +

Evolved High Speed Packet Access is an upgrade for UMTS that is a direct evolution from the HSPA protocols. It was standardized by 3GPP in release 7 and continued until release 10. HSPA+ increases considerably the data rate speeds, being comparable and competitive with new LTE networks, in fact, its network architecture has several similarities with LTE networks; but without needing to deploy a it new radio interface. This allows the operators to continue to use the HSPA network and has speeds competitive with the 4G technology. The theoretical values for data rates are 168 Mbps for downlink and 22 Mbps for uplink connections. These values are achieved due to the use of multiple-antenna technique (MIMO, multiple inputs multiple outputs), a better modulation scheme, 64QAM and a technique that allows to gather multiple cells into one (Dual-Cell HSDPA). One option for the HSPA+ is a all-IP based network that is an important objective of mobile telecommunications network, because it reduces the cost per megabyte of data, improve the latency and simplifies the connection with the broadband. The range of advantages described for HSPA+ improves the performance of real-time conversation and interactive services and it also allows the use of VoIP and the use of heavier multimedia content.

2.2.2 Long Term Evolution

Long Term Evolution is a standard for mobile telecommunications, usually described as 4G that achieve the highest data rates available in the market. Its specification was completed for the 8th release of 3GPP in the beginning of 2009. LTE is an evolution of GSM and UMTS standards, with a different radio access network and several improvements to the core network. Since it is descendant from these technologies, LTE can coexist with both 2G and 3G networks being even possible to have the three technologies available at the same place.

LTE uses Orthogonal Frequency Division Multiple Access (OFDMA) that is a multiplex technique for the downlink connections; and single carrier FDMA for the uplink connections, these techniques allow an efficient use of spectrum. These techniques are different than ones the used in 2G and 3G; therefore a new air interface was needed and was called E-UTRA. LTE uses a full IP network architecture contrarily to its predecessor that used a circuit+packet switching network. With all the features described, LTE is able to deliver a download data rates of 300 Mbps and upload data rates of 75.4 Mbps superseding all the previous technologies. It has also lower data transfer latencies, a faster handover and a faster connection setup time compared to the previous technologies. It can also support communications in mobility scenarios being able to support terminals moving at a speed of 320km/h.

2.3 3GPP Network Architecture

The 3rd Generation Partnership Project is an organization composed by six telecommunications associations (ARIB, ATIS, CCSA, ETSI, TTA, TTC) that aim to develop and standard the specifications for 3GPP technologies. The initial objective of 3GPP was to make a standardization of the implementation of 3G networks based on the previous technology, the GSM, created by ITU. After that, the main objective of the project was extended to support, maintain and further develop the radio access network, core network and service architecture.

As stated before, the group operates in several areas of telecommunications network technologies, starting by the radio access network up to the core network. It specifies the all architecture of the radio access network, from the radio capabilities of the phone to the access base stations and all the modules from the core network (in charge of users charging, network policies among others) are also specified. Service capabilities like security of the network, codecs development and quality of service are also some of the services developed by 3GPP.

In this section, it will be described all the architecture of the network specified by 3GPP, with elements from different network modules like LTE, EPC, PCC and also radio access technologies [1]. In figure 2.1 it is possible to overview the network architecture.

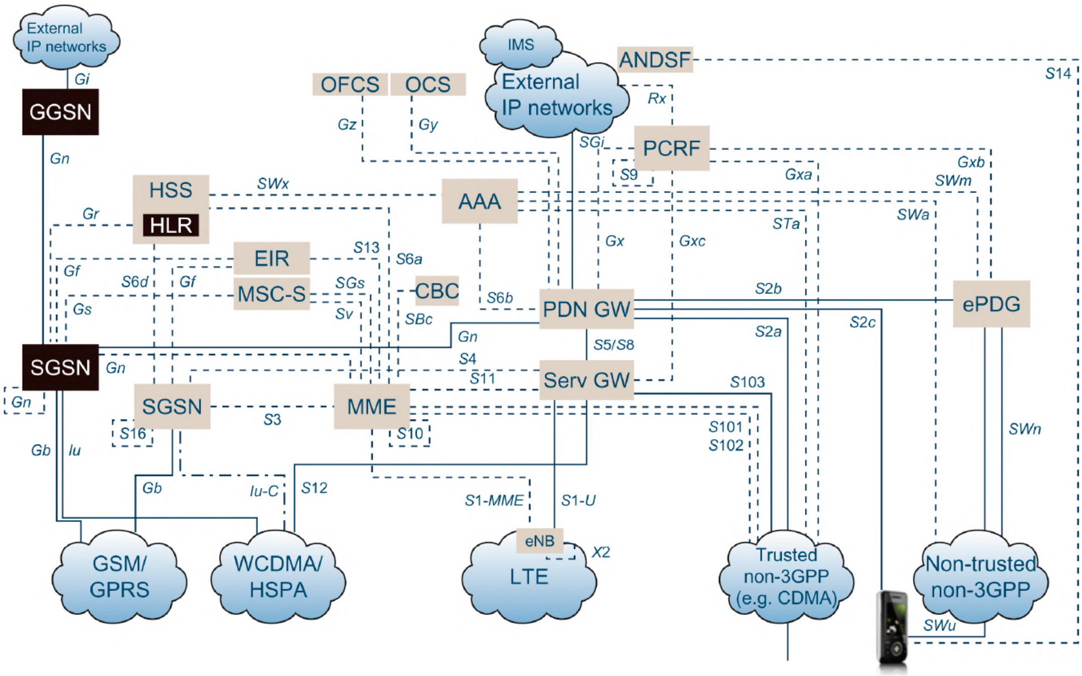

Figure 2.1: Network Architecture [1]

The next section details the elements that belong to the policy and charging control.

2.3.1 Policy and Charging Control

Policy and Charging Control (PCC) is a module from the core network that assembles functionalities of the previous 3GPP network releases; it is a centralized solution with the main objective of guaranteeing that IP sessions have the proper bandwidth and QoS, but is also capable to deliver services to control charging per subscriber.

This solution enables network providers with the capability of creating new ways of revenue, by guaranteeing bandwidth and QoS on specific services, and also by being able to charge for specific services among other types of charging policies. PCC has also the capability of dynamically controlling the network resources and subscribers, using pre-defined policies and also dynamic PCC rules. These rules are constructed based in the network information like QoS parameters, type of service, charging policies and monitoring factors.

In fact, PCC is an assemble of several components that work together to give the network these capacities. The architecture of PCC is in figure 2.2 and the components that belong to PCC are the following [2]:

- Policy and Charging Rules Function (PCRF) it a module with capacity to control data flows by the subscriber and to create rules based on charging (described in detail in 2.3.1.5);

- Policy and Charging Enforcement Function (PCEF) is implemented in the serving gateway and it is able to enforce the rules in the network to be executed;

- Online Charging System (OCS) is a module that controls the online charging system by giving credit management and credit information to PCEF based on different chargeable events (data transferred thresholds, specific services, time of connection);

- Off-line Charging System (OFCS) is a module that records data concerning charging to be stored for the billing system;

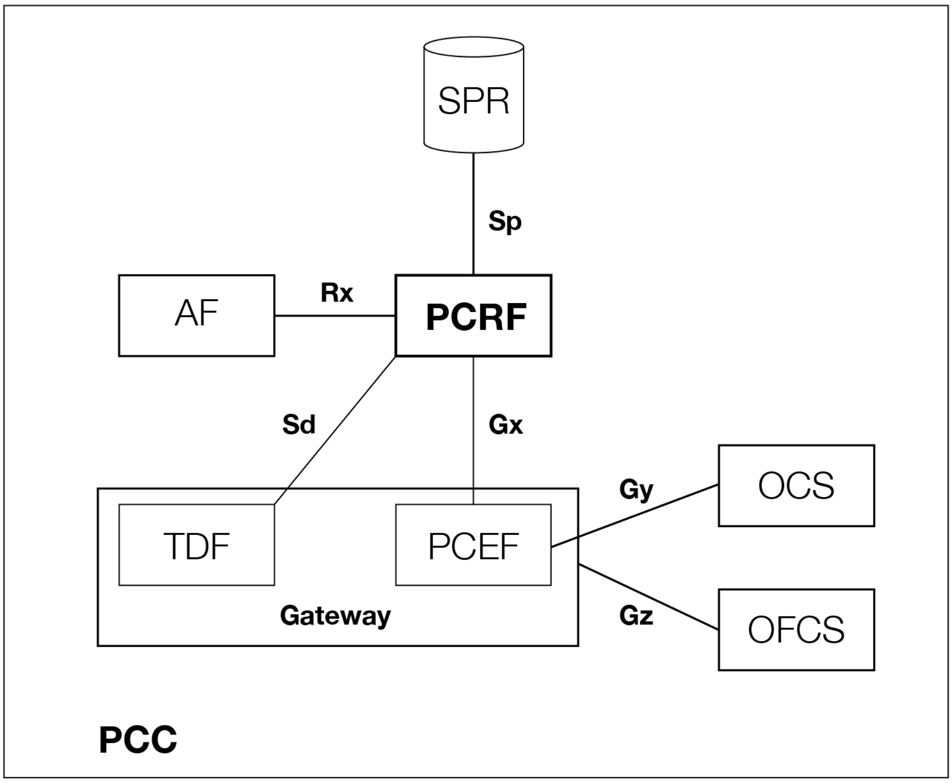

Figure 2.2: PCC architecture

2.3.1.1 Subscriber Profile Repository

As the name of the module indicates, its main function is to keep the subscriber profile's information. It contains useful information to create rules based on the user subscription type, it allows services and charging information This repository can be accessed by PCRF through the Sp connection to be able to construct its rules with the information mentioned above.

2.3.1.2 Application Function

The AF is a component of PCC that interacts with services and applications being able to obtain information useful to PCC dynamic Rules. For example, AF is capable of extracting information from sessions using application signaling and giving details like Application Identifier, bandwidth, subscriber identifier, flow status, etc. The information acquired is then provided to PCRF by the Rx connection to be used in its rules.

2.3.1.3 Traffic Detection Function

Traffic Detection Function (TDF) is a module in the core network that works with the principal objective of detecting service data. It can perform actions in services' traffic like traffic shaping, redirecting, blocking or permite its use without restrictions (working with PCEF). The information gathered by TDF, similarly to other modules, is shared with PCRF, using Sd reference point, to be included in its policies/rules. They do a combined work to impose the rules to the network and have its effects in users services.

2.3.1.4 Policy and Charging Enforcement Function

The Policy and Charging Enforcement Function (PCEF) is the module in PCC that enforces the rules in the network to be executed. The policies and rules come from PCRF; these two modules change Diameter [17] messages between them over the Gx interface. Besides having the task to enforce the effects of the rules in the network (install, modify or remove policies), PCEF is also capable of detecting changes in service data flows and it is able to report them to PCRF (this is possible due to the PCEF location, at the gateway).

PCEF is also capable to work without PCRF, and it supports data flow detection, charging depending on service and it also does the enforcement of pre-defined rules. To do this, PCEF is able to interact with online charging system (OCS) using Gy connection, and report the resources usage to offline charging system (OFCS) using Gz interface.

2.3.1.5 Policy and Charging Rules Function

Policy and Charging Rules Function (PCRF) is a module of Evolved Packet Core with the main objective of applying rules to the network to impose flow control and charging to subscribers. It was introduced in 2007 when the PCC architecture was released and contains multiple solutions of PCRF from different vendors. Despite the fact that some of its functions are present in previous versions of networks specified by 3GPP, they were never as centralized as in PCRF. This fact permites its access to subscriber data bases and the use of specific functions to evaluate which rules it must use. Due to centralization, PCRF can operate upon multiple wireless technologies.

One of the main scope of PCRF is the user charging system, that due to its capacities, it became very versatile with various solutions to impose charging to users (example of controlling the transferred data to charge it). Other important feature of PCRF is its granularity, as it can make specific rules to each user in the network. With this type of solution and combining it with other network elements, such as service monitoring systems, and knowing that it can take its decisions in real-time, PCRF is a powerful tool in every network. It allows the creation of new solutions to operators like dynamic allocation of bandwidth depending on various factors (time of the day, application used) or control and guarantee QoS for determined services and for different types of subscribers.

2.3.2 Universal Terrestrial Radio Access Network

UTRAN is the abbreviation for Universal Terrestrial Radio Access Network and it is the fixed network infrastructure that makes the intermediate between user (with an user equipment) and the core network. This radio access network is used in UMTS suing a W-CDMA air interface. 3GPP specifies its architecture, the physical layer and the radio technology.

Besides the fact that the main access technology is W-CDMA, UTRAN also support other technologies like Wi-Fi and the radio access technology of the previous mobile technology GSM/EDGE.

The modules that composed the UTRAN are the NodeBs and the RNCs and there are four types of connections associated with UTRAN: Uu, IuB, IuR, IuPS and IuCS. Uu is the connection between user equipment and the NodeB, the IuB connection is established between the NodeB and RNC, the IuR connects different RNCs, and finally, IuPS and IuCS are the connections between UTRAN (that is a radio access network) and the core network, for packet switching and circuit switching respectively. The architecture is described in figure 2.3 [3].

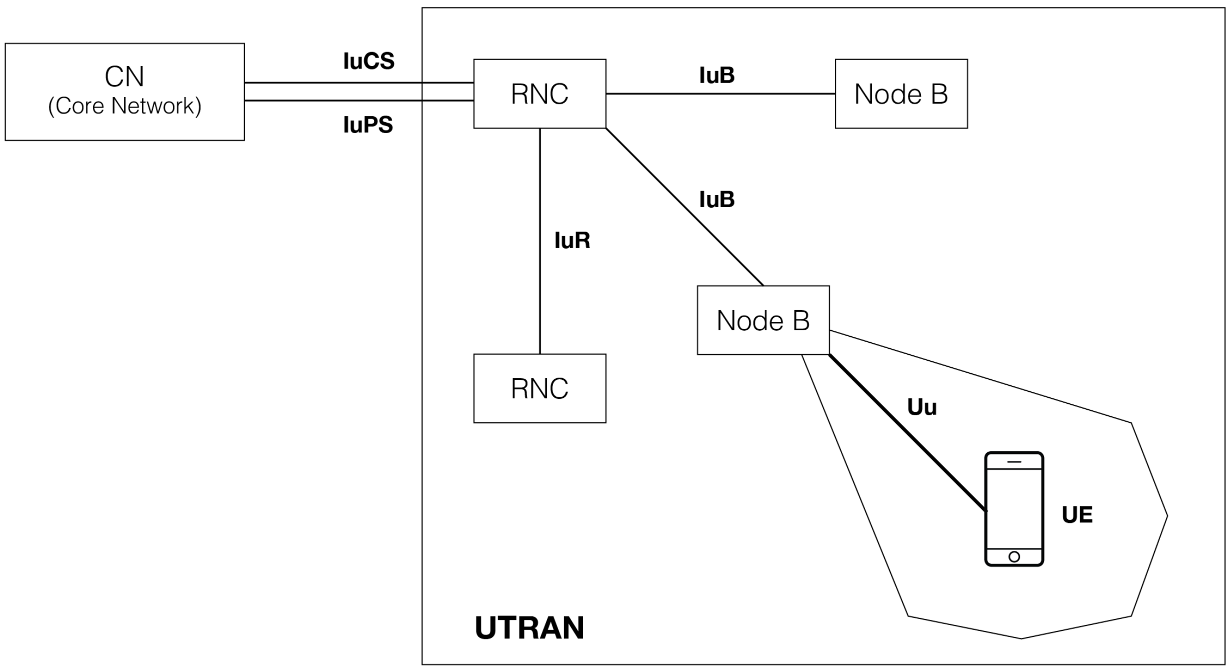

Figure 2.3: UTRAN

2.3.2.1 User Equipment

User equipment is a device that is used by the user/subscriber, and it can communicate with a NodeB or E-NodeB in UMTS and LTE networks respectively. This device is equipped with an antenna that is in charge of doing the RF connection to the NodeBs, called Uu in UTRAN. Nowadays the user equipment can be a smartphone, a 3G USB modem or any equipment that can access the mobile telecommunication networks. The user equipment has functions like:

- Radio transmission initiation and termination;
- Call control;
- Authentication;
- Session management;
- Mobility management.

2.3.2.2 NodeB

NodeB is the name of the equivalent for base transceiver station in GSM, but now in UMTS. It is a component of the UTRAN that is controlled by a RNC, and it is the one that makes the connection from the user equipment with the network. The basic concept of NodeB is the same as of BTS: it contains the equipment for transmitting and receiving radio signals to and from the UE's, and it is in charging of create the radio link conditions to do so. W-CDMA is the technology used to create the air interface between NodeB and the UE. In this configuration the UEs can move freely around the nodeB and communicate with it; communication between UEs is always made through the NodeB (they do not connect to each other directly). Physically, nodeB is composed by an antenna (or more), digital signal processors, power amplifiers and backup batteries for the case of power failure.

2.3.2.3 Radio Network Controller

The Radio Network Controller(RNC) is a component of UTRAN that is in charge of controlling the NodeBs that are connected to the RNC. The basic functions of RNC are related with management of the cells resources like handover, power control and channel allocation. RNC is also the component which makes the connection between UTRAN and the core network: using IuPS connection, it reaches SGSN that belongs to the Packet Switched Core Network, and with IuCS it connects to MGW that is part of Circuit Switched Core Network. To communicate with the different modules that are connect to the RNC, there are different protocols that are implemented in the RNC [54]

- NBAP (NodeB Apllication Part) is used to communicate control of the NodeBs connected to RNC;
- ALCAP (Access Link Control Application Protocol) is used in the control plane for the transport layer with the basic function of multiplexing different users in one AAL2 transmission path. It is used in the IuCS connection but also in the NodeB on IuB connection;

- RANAP (Radio Acces Network Application Part) is the protocol used in Iu interface that connects the RNC to the core networks (to the MSC and the SGSN);

- RNSAP (Radio Network Subsystem Application Part) is used to make communications between RNCs (over the IuR) connection.

The RNC is the main component of the radio access network; therefore the principal logic of the UTRAN is in the RNC. The main functions of the RNC are:

- Call admission control: the RNC is capable of deciding if it accepts a call or if it rejects it depending on the interference in the cell;

- Protocol conversion: the RNC is able to connect the core networks, the NodeBs and also another RNCs; to interchange messages between them is necessary the conversion of protocols;

- Handover: the RNC is in charge of controling the handover process and the signalling needed to do perform the handover;

- S-RNS Relocation: if the UE gets out of the scope of one RNS, it is necessary that another RNC assumes the control of the connection made;

- Codification: the data received from the core network must be coded to deliver to the user, and the data received from the user must be decoded before it reaches the core network;

- O&M: the RNC provides data related with the radio access network status and evaluation (metrics and KPIs) that can be used by the operators to control the network;

- Radio Resource Control: the RNC manages the radio resources of the NodeBs and plans the channel utilization;

- Code Allocation: since UMTS uses W-CDMA technology in its air interface, the RNC decides which codes are available and allocated to each connection;

- Radio bearer set-up and release: each connection is called a bearer and RNC is in charge of set-up, maintain and release these connections;

- Power Control: the RNC controls the power used by the NodeBs in their connections;

- Congestion Control: in case of congestion, it takes actions to resolve it, stop accepting calls and drop connections are examples of actions to be taken.

2.4 Congestion in Cellular Networks

Due to the evolution of technology, it is now possible to have more services and applications in every UE. This growth of bandwidth use must be accompanied by a growth of network capacity, and sometimes it is outpaced by the demand of resources. Since the RAN provides the basic link between clients and the network, as the RAN becomes congested, the users start to have problems in theirs connections and the number of dropped connections increase dramatically. Although it is hard do define W-CDMA capacity due to its characteristic of having soft capacity (the operator can define if it wants to support more clients and reduce the quality of the service or the vice versa, contrary to 2G that is constrained by TRX capacity), it is possible to access some of the causes of congestion, that are described in subsection 2.4.1, 2.4.2, 2.4.3 and 2.4.4 [34].

2.4.1 Power

Concerning the physical layer, the power necessary to maintain a communication between NodeB and UE may vary depending on factors like distance to the NodeB or trying to maintain a good CIR (Carrier-to-interference ratio). The increase of users in a cell also increase the necessary power to maintain all connections. Therefore, sometimes the power available on the cell is not sufficient to keep and accept new connections; when this happens, it is considered that it is occurring power congestion [31].

2.4.2 Channel Element

Channel Element (CE) is defined by the capacity and hardware required by a NodeB to make a connection for one Speech Radio Access Bearer(RAB) (with the codec Adaptive Multi-Rate(AMR) with a bitrate of 12.2 kb/s). The definition of channel element is not standardized by 3GPP, so it can vary depending on the vendor of the NodeB (some vendors include in the definition of CE the control plane signalling necessary to do the connection for example). The relation between channel elements and bandwidth is a direct relation (the higher the bandwidth the higher the number of CE required), however this relation could not be a proportional [25].

SF	CE ladder	Service (Example)
256	1	AMR4.75 AMR5.9
128	1	SRB(Stand-alone)13.6 AMR7.95 AMR12.2 AMR12.2+PS0 AMR12.65 PS16
64	1	
32	2	CS57 CS64 CS64+PS8 PS64 AMR12.2+PS64 Streaming 64+PS8
16	4	PS128 Streaming 128+PS8
8	8	PS384

Figure 2.4: Channel elements relation with bandwidth for Ericsson [25]

2.4.3 Code Tree

UMTS uses W-CDMA as an air interface to stablish communication between the NodeB and the users in that cell. As the name indicates, W-CDMA stands for Wideband Code Division Multiple Access, a technique that assigns different codes to the users to make a connection with the NodeB. These codes are finite resource of the NodeB; therefore, they can be all required, and consequently, the starvation of these resources can happen [52]. There are different types of codes in the NodeB, both short and long codes. The architecture of the code tree can be seen in figure 2.5. Long codes are the ones at the right of the figure and are the ones with a higher spreading factor. This is the opposite to the short codes, that are at the left end of the picture and are associated with a lowest spreading code. The use of a short code implicates the blocking of several long codes. Since the length of the code is equal to the length of one symbol in chips, it is possible to conclude that the short codes represent a higher bitrate than the long codes. In fact, long codes are used to make connections requiring lesser bandwidth (the NodeB is capable of doing a lot of these types of connection), and the short codes are used to connections requiring higher bandwidths.

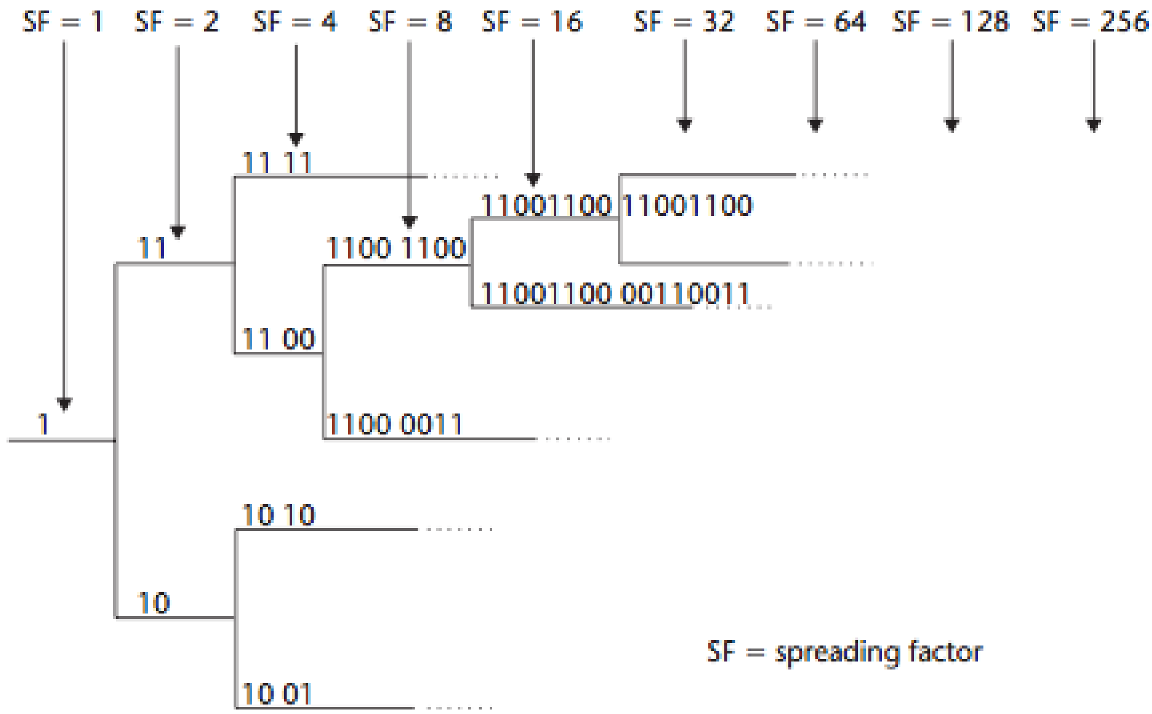

Figure 2.5: Code Tree Channelization [54]

2.4.4 Bandwidth in Iub

The Iub connection (can be seen in figure 2.3) connects the NodeB to the RNC. This connection can be the cause of congestion in 3G due to its bandwidth limitations: the often increase in capacity of the air interface outpace the Iub connection capacity. Associated with a high cost that demotes the operator to improve them, these links are usually subject of congestion.

The work in [43] showns that the evolution of 3G technologies, especially with HSDPA, increased significantly the amount of bandwidth available in the RAN to users. This increase represent a problem to RAN because this evolution was not followed by an increase of the capacity of Iub connection (the connection between NodeB and RNC). The high cost to update this connection is the main reason to its lack of improvement. To avoid this situation, it is proposed the use of an algorithm that provides de congestion control needed to avoid congestion in Iub. This work shows that the congestion detection techniques introduced can keep the level of utilization of Iub bandwidth high with low delay and loss.

2.5 Metrics

This section describes the RAN metrics, which describe the state of the radio access network. With the right evaluation of these metrics, it is possible to achieve rich information on the state of the RAN. Levels of congestion or occupancy are some of the main topics that can be obtained with metrics. This section describes in detail the types of metrics given by network equipments, and it also describes some network management systems that gather these metrics and make them available to manage the network. To determine the status of the RAN, it is possible to check a series of network parameters. At the lowest level, the network is composed by networking equipment that can be accessed and monitored. The parameters given by the network are Key Performance Indicators (KPI), Performance Indicators (PI) and counters, that describe a wide range of actions and events that happen in the RAN. They can be accessed directly on the equipment or by their Element Management Systems (EMS).These metrics are very useful to telecommunication providers due to its detail of the RAN status; with them, it is possible to find problems in the network/equipments, cells that need to have theirs capacity improved or even detect congestion. Since there is a wide range of radio access network equipment providers, the description of metrics may vary as well as the frequency of their update, although they seem to be very similar from vendor to vendor.

As stated before, the components of the RAN can be constructed by a wide range of manufacturers, but they seem to cover the same type of network parameters. These parameters cover several areas of RAN: Transmission Network parameters, Equipment parameters and Radio Network parameters. The first area is related with the signalling necessary in the network and the protocols used. The second area is related of parameters of the physical equipments performance (e.g. CPU usage, processor modules, etc). The final area describes specific events in the network related with radio. This last area has several subsections that describe in detail the performance of the radio part of the network:

- Accessibility: provides the accessibility of the cell like PIs of attempted call setup and sucessful;
- Retainanbility: gives statistics of cell call drops with deep specification of the reason and type of connection drop (RAB release for PS domain, CS domain etc);
- Mobility: provides metrics about handover performance, from incoming handovers to outgoing handovers (inter technology or intra technology), number of successful and failed handovers, number of attempts and number of users trying handover;
- Traffic Volume: provides information on the transferred data for different types of technologies (HSPA, r99, etc) with the specification for uplink and downlink and even for different classes of service (stream, interactive, conversational and background):
- Resource and Load Indication: this information is the most related with congestion, it shows network resources occupancy and fails related it them. The resources described in sections 2.5(Codes), 2.4.1 (Power resource), 2.4.2 (Channel elements) and

2.4.4(bandwidth in Iub connection) have here parameters that describe their performance. These metrics also give information on the number of users connected to the cell.

- Uu, Iub, Iur, Iu Standard Signalling: provides metrics on the signalling necessary to keep RAN connections working;

With the description of the fields shown, it is possible to monitor the access network and give useful information for the creation of PCC rules [29]. Some of the metrics may need to be combined with others to give advantageous information [32].

2.5.1 Network Management Systems

This section will describe some network management systems available in the market by telecommunications vendors, witch will gather the metrics described before. The AriesoGeo system is a product that gives metrics and KPIs geographically located within the network representing an innovative solution. The Altaia platform is PT Inovação product that gathers metrics from the RAN, and it is going to be used as a component in the proposed solution of this Dissertation.

2.5.1.1 AriesoGeo

Arieso Geo [15] is a product developed by Arieso that consists in a platform able to locate, store and analyse data coming from radio access networks, to deliver rich information to operators. With ariesoGeo platform, Arieso states that it is capable of delivering precise information on users location inside the cell. Besides of this information, ariesoGeo is also able to recognize problems in the radio access network, for each user, being capable of locating the user and provide its true experience (this information is provided without the use of GPS). AriesoGeo reduces the OPEX to the operator, providing reliable information on locations with problems dispensing the use of manual methods to identify the problems in the network. This product provides maps with the operator infrastructure and also shows where the problems are occurring, it is a great tool to identify areas that need the capacity improved and the detection of congestion/failures in the access network. AriesoGeo is prepared to work with different equipment vendors and with different access technologies, being highly adaptable and versatile. AriesoGeo delivers geographical KPI's (like setup connection failures due to congestion and drop RAB calls) for the network and can even configure alarms with thresholds by cell. It can also provide detailed information on the consumer behaviour allowing the operators to explore new opportunities of business.

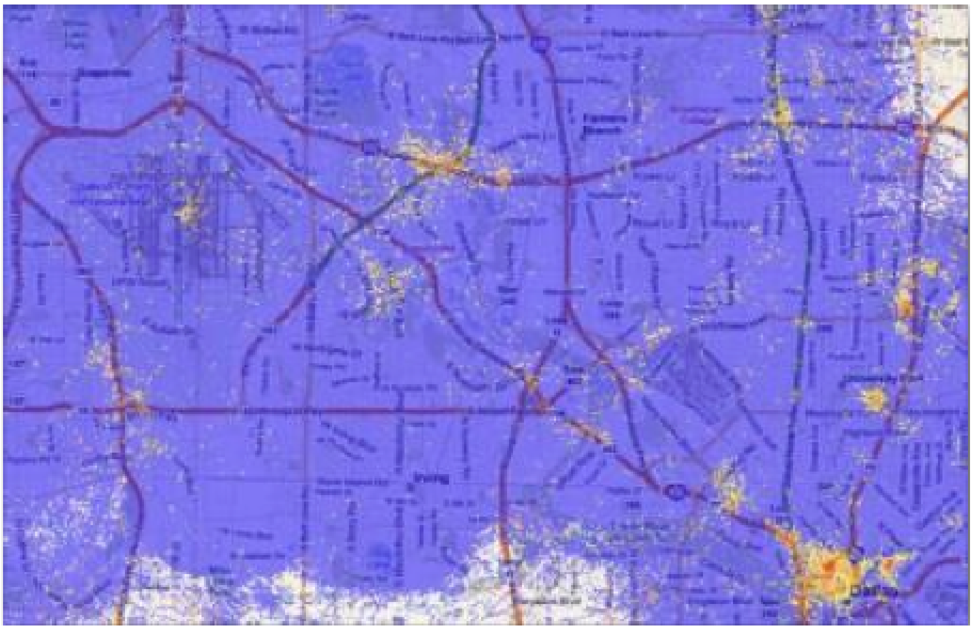

Figure 2.6: AriesoGeo Figure with KPI disposed geographically [15]

2.5.1.2 Altaia

Altaia is a platform developed by PTinovação that has the main objective of storing KPI's and QoS parameters of the network. It is also capable of delivering statistics of the network performance, being very useful to service providers. It uses several modules to achieve this type of monitoring:

- CDR and meter analysis: it is capable of acquire information related with charging and do its analysis;
- traffic measures and network performance: measures the traffic done with different access technologies and evaluates the network performance;
- quality of services usage measures: gathers information about QoS of the services provided by the network;
- service guarantee analysis: analyses the availability of the services in the network, like down times;
- threshold alarms generation: based on the comparison of metrics with thresholds, Altaia is capable of generating alarms;
- performance alarm monitoring and processing: considering metrics that evaluate the performance of the network, it can monitoring them and process it information;
- network and service metrics management (KPIs and KQIs): Altaia can manage metrics from different vendors achieving the principal KPIs and KQIs used to evaluate the status of network and allowing the service providers to manage their networks;

- SLA (Service Level Agreement) management: evaluation of metrics related with service availability and response time.

With the information gathered, Altaia provides the capability to discover network problems and give information on the cells that need to be upgraded. The information can also be used to construct network policies since it can give information on congestion status for example.

As stated before, Altaia is capable of delivering metrics about the radio access network status, like key performance indicators on the radio resources of cells. These metrics belong to different types of areas that describe the network, areas like the ones described in section 2.5. The main advantage of using the Altaia platform is the feature of assembling metrics from network equipments (like RNCs or NodeBs) from different vendors in a standardized mode; despite the different names and modes of representing them, Altaia maps the native vendors' metrics in a standard mode developed by Altaia, enabling the network providers to access all metrics information of their network in a single platform display equally to all cells (regardless of their manufacturer).

In this Dissertation, Altaia is going to provide the metrics that show the state of the radio access network, to be possible to evaluate the occupancy and congestion in the RAN. The metrics provided are mainly related with the factors of congestion in cellular networks described in section 2.4, the channel elements used and its faults, the codes resource congestion and the load in the cell. These informations will feed a module that is capable of evaluate them into a general state of congestion for the specific cell,s and report it to PCRF, so it can actuate in the network with congestion resolution policies [47].

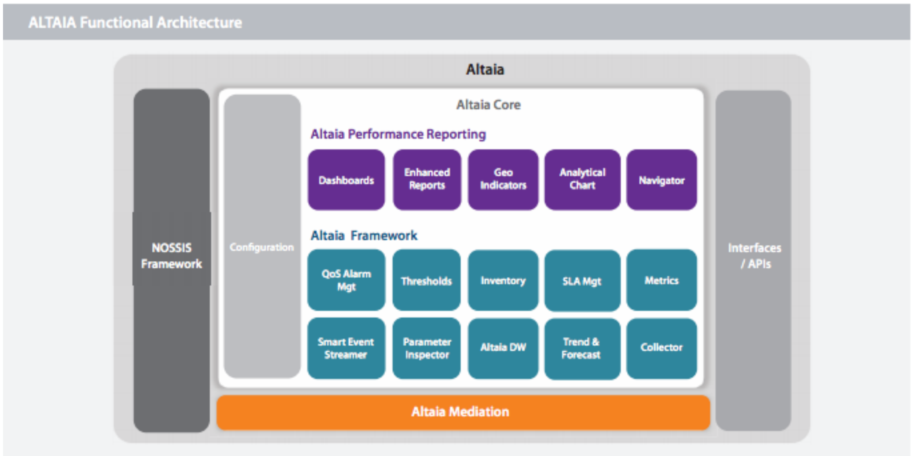

Figure 2.7: Altaia Architecture [47]

2.6 Probing

Passive probing is used in telecommunications environment to acquire information that enables the a monitoring of the network. The detail of the monitoring depends on the information gathered by the probe, that can be user specific or QoS oriented. Concerning probing on the user data, it can bring information to network providers that enables them to do statistical analyses of user behaviour (the traffic volume, duration of sessions), and probing on applications enables providers to know which are the applications that are used more and the traffic demands for them.

In this Dissertation context, probing as an important rule since it enables the acquisition of important data, mainly related with user behaviour in the network (bandwidth usage per example) and also specific application content (the applications used and the amount of data transferred in it). This information can be later used to integrate in policies to maintain users QoS. In this section, we will mainly give information about probing in streaming applications, since these will be the ones we will work in this Dissertation, and about Deep Packet inspection (DPI), which we will use to aquire services information.

2.6.1 Streaming Applications

Streaming became very popular, and due to that, it has increased exponentially the amount of traffic on the operators' networks. However, video streaming is very sensitive to delay and requires a high bandwidth to operate flawlessly, so to keep a good QoS level it is very important to have it monitored recurring to probes.

2.6.1.1 Meo Go

Meo Go is a service created by PT that offers television channels over the internet with a high acceptance by the subscribers. The service is available for multiple platforms like smartphones, tablets and computers. Meo Go also supports the visualization of movies from the service Video Clube [38]. It has also other features like TV guide to know every scheduling of the TV contents or schedule recording of TV content. It uses a smooth streaming player of Microsoft to deliver an adaptive stream service using HTTP. Due to the architecture used to deliver the service, it is possible to achieve parameters that show its is behaviour and problems, being able to determine a QoS measurement, recurring to a probe.

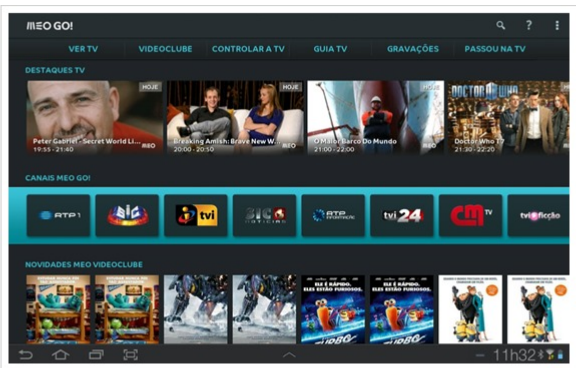

Figure 2.8: Meo Go

2.6.1.2 Youtube

Youtube is another phenomena of popularity for video streaming visualization. It uses an adaptive streaming technique from Adobe called Adobe Dynamic Streaming for Flash [5]. It is very used in mobile networks mainly due to its applications, that are spread across multiple user equipments. These factors have has consequence the generation of high amounts of data. Due to this fact, and also considering that the flawless play of this type of video requires excellent conditions from the network (like high bandwidth and low latency), this is an application that must be probed to guarantee a good quality of service.

2.6.2 ArQos

ArQos [27] system platform is a PT inovação product that evaluates QoS and QoE in networks. Using different probing methods, ArQos is able to achieve information on the perceived network, service quality and performance level experienced per user and per service. This solution allows service providers to assemble information on the quality experienced in their networks, and therefore optimize and troubleshoot problems that appear. It is able to perform probing in landlines and mobile networks being able to benchmark these networks. This solution may also be used to test new services or upgrades made in the network before their commercial release. To achieve it, ArQos uses several probing systems divided in two categories [18]:

- intrusive probing - also called active mode, uses both hardware and software components. This type of probe interferes in the normal functioning of the networks to infer the quality of services that are being provided. They are used to measure the service availability, quality of service on radio and network performance simulating typical user's activity and events in the networks.

- non intrusive probing - also called passive mode, has hardware and software components that do not interfere in the normal functioning of the network and achieve KPI's of its functioning. It analyses traffic in the network like VoIP traffic including signalling messages to evaluate QoS and QoE.

The type of probes and some products available in ArQos environment can be seen in figure 2.9.

The next section shows the ArQoS NI LTE/EPC probe with special interest to this Dissertation, since it is capable of delivering KPIs on the state of the RAN and the users.

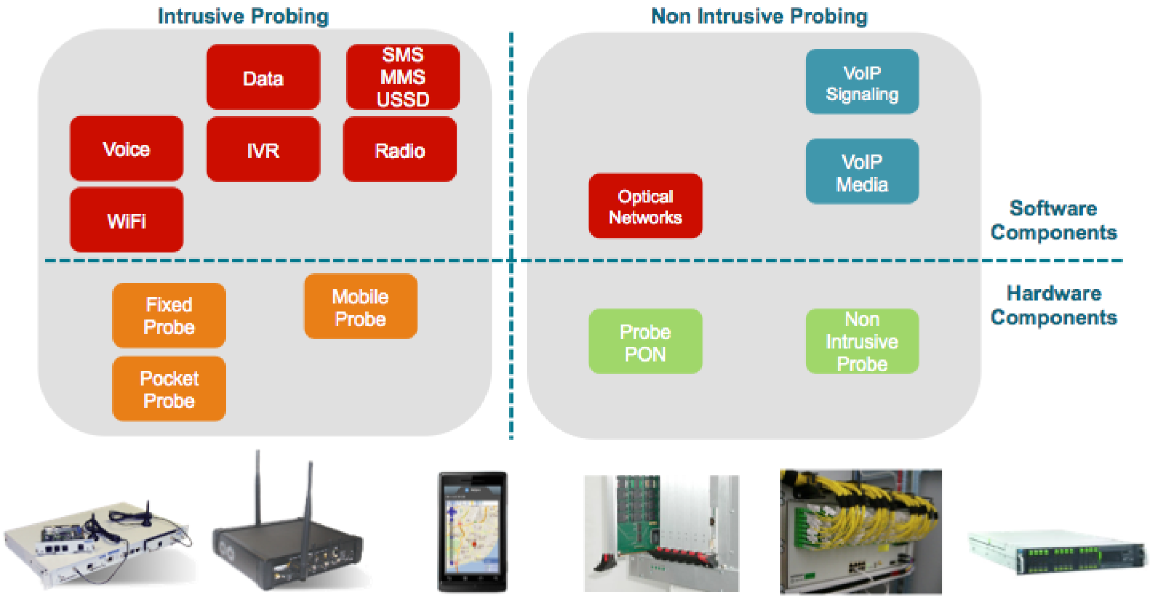

Figure 2.9: ArQos

2.6.2.1 ArQoS NI

The ArQos NI is a passive probing solution developed in the ArQos environment for LTE/EPC. It collects network traffic and analyses all VoIP calls permitting integration with operations support systems. This probe is able to detect attachments, detachments and handovers events in the RAN. The information gathered by this probe can be divided in six categories:

- Accessibility: provides data on the attachment and detachment of users with the E-NodeB like the number of attempts, number of successes and number of failures;

- Mobility: gives information on handovers done in the cell; the preparation failure rate, completed/successful handover rate, detection of incoming handover and delay, information on inter-technology handovers (like 4G to 3G or 4G to 2G).

- Retainability: number of released connections (ex call drop) with cell identification;

- Operation&Management: data volume transferred in RAN and Mobility Management Entity (MME) overload;

- Network usage, quality and integrity: gives the throughputs proceeding in cell identified by IP, E-RAB or user profile;

In figure 2.10 it is possible to see the positioning of the ArQoS NI probe in the LTE/EPC network.

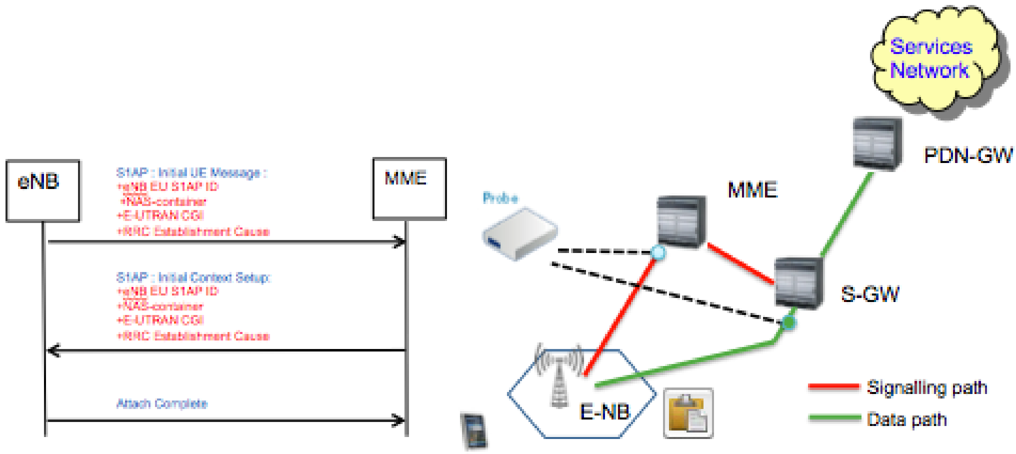

Figure 2.10: ArQoS NI probe

2.6.3 Deep Packet Inspection

Deep packet inspection (DPI) is a technique used in networks that allows the inspection of packets that are in the network. This action permits obtaining high level information (at layer 7) that can have different usages [13]. For network providers, DPI can have various utilizations. One of them is to acquire QoS of services by inspection of their packets, and take actions in the network according to the information on the packets. For example, information that shows that the user is having a video conference call are good indicators to the operator to increase the bandwidth of this client, to ensure that the video call is made with a good quality.

On the market there are several DPIs from different vendors. The LanGuardian [44] DPI solution can perform network traffic monitoring and detect congestion. It can also detect users that are receiving live stream services. The Dynamic Actionable Recognition Technology (DART) [11] is a DPI from Allot that can do application, device, user and network topology identification. It can also provide visibility of the network traffic and discover QoS of services to deliver better QoE to the users.
In this section, it will be described in detail several more DPI solutions with particular interest to this Dissertation, due to its characteristics like application identification and user consume profile and QoS.

2.6.3.1 Qosmos Deepflow

Qosmos Deepflow [51] is a DPI probe software for telecommunications that is able to collect data from real time traffic flows and extrapolate rich information to third-party systems. This probing system is passive, which means that it does not interfer in the normal flow when it is measuring. Qosmos Deepflow probe is able to work in two different configurations, to achieve two types of specified information:

- Subscriber Analytics [51]: in this mode the probe functioning can provide specific and detailed information, user oriented, ideal for creation of user profiles and preferences (data aggregation per minute, per user and per application). This probe configured in this mode does a layer 2-7 IP flow analyses, using DPI techniques and also statistical and behavioural analyses. It can work on wired lines and also in celular networks. This probe can achieve very detailed information on the user, it can detect encrypted applications like P2P or skype, identify applications embedded in Facebook, identify different types of service like chat, file transfer or VoIP done by applications like skype or Instant Messaging. This probe can also detect metadata from applications like HTTP URL, Facebook user profile or even relations in Twitter, its architecture can be seen in figure 2.11.

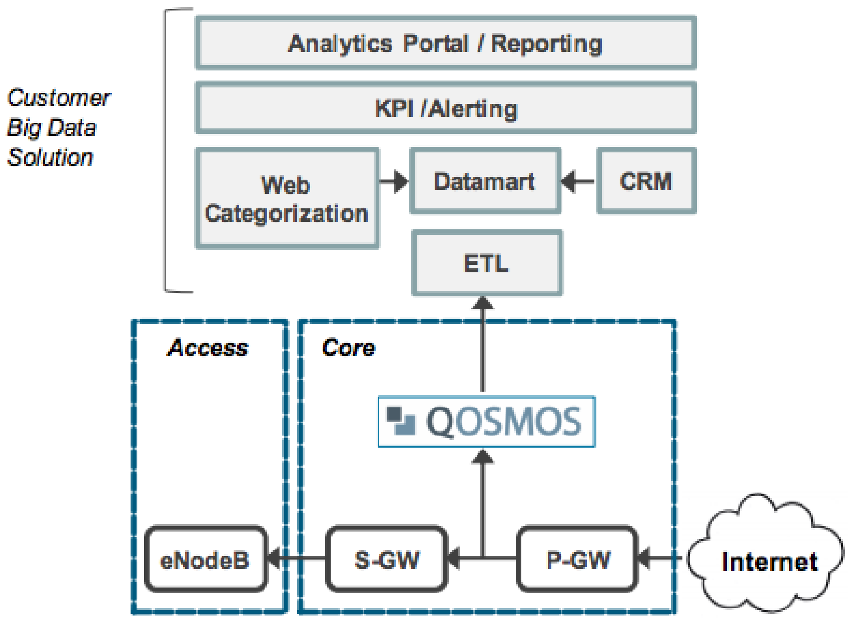

Figure 2.11: Qosmos User [51]

- QoE [50]: In this mode the probe works to give information on the QoE of each user. Once again, it analyses the layer 7 traffic using DPI, to acquire information on application and services and posteriorly acquire metrics that evaluate the QoE. This probe can give information related with delay, jitter, volume in bytes and session duration that are very useful to determine QoE. It can extract TCP metrics like application response time, packet loss and disordered packets and in HTTP it can achieve the object download time and return codes. With all these parameters, Qosmos is able to do a detailed report including the user identification, the applications used and its QoE. Its architecture can be seen in figure 2.12

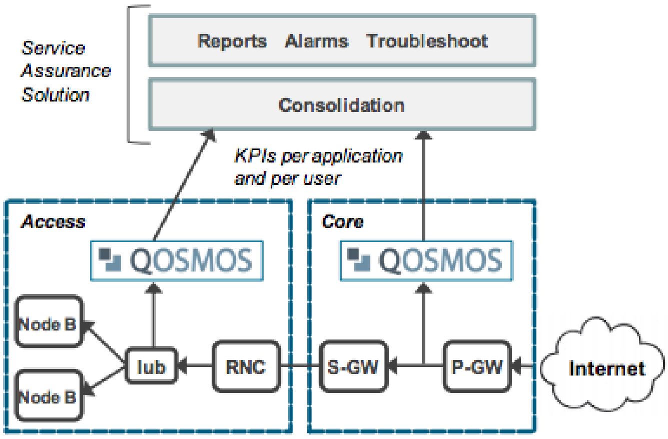

Figure 2.12: Qosmos QoE [50]

2.6.3.2 Procera NAVL

The product NAVL [49] of Procera is a DPI that provides real-time, layer-7 application classification and metadata extraction for network. NAVL applies techniques of packet inspection, allowing pattern matching and recognizing the type of service. This method of pattern matching allows the creation of behavioural and statistical analysis that can be used to implement policies in the network. With the service, recognizing it is also possible to achieve specific metrics of that service and in some cases achieve its QoS. The NAVL product can do identification and metadata extraction from applications like YouTube, Skype and Outlook. In figure 2.13 is possible to see its architecture and functionalities.

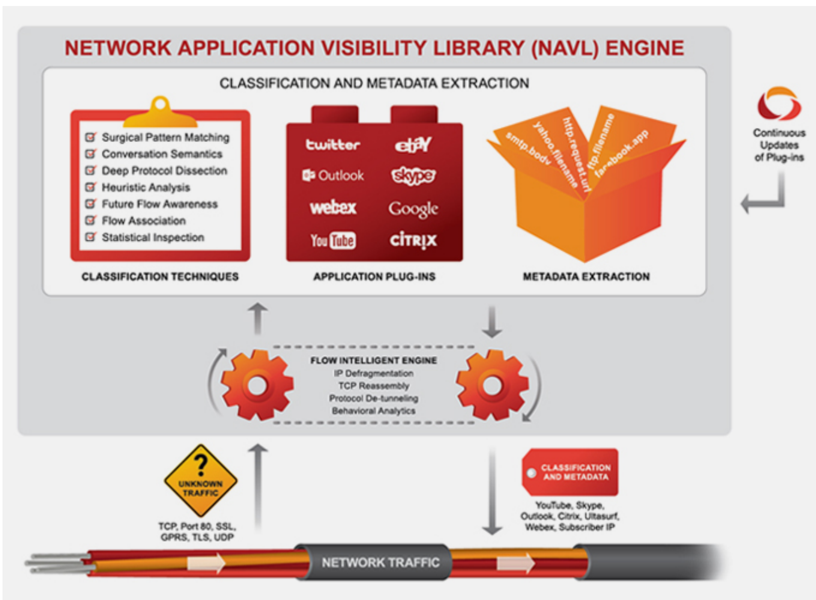

Figure 2.13: Procera NAVL description [49]

2.7 Call Admission Control

Call admission control (CAC) is a module used to deliver policies into the network with the main objective to control admission of sessions and the congestion of the network. The principal component to create these policies is PCRF (that was described in section 2.3.1.5), that can implement dynamic rules to control the network resources in order to keep it working without problems and at the same time keep the profitable.

Considering the increase of user bandwidth consumption, the problems of the network are mainly due to the lack of resources, which often cause congestion. To guarantee that the users service is done with the best conditions, it is necessary to infer the status of the network and also the QoS [24]. Call admission control policies are then based on elements (sections 2.5 and 2.6) capable of delivering this information on the networks status. QoS can therefore be obtained using methods like probing, and new solutions can be developed based on QoS and network status [21].

The work in [40] presents a solution to improve the usage of the network resources, based in policies using a module that can acquire information about the occurrence of congestion. The provided information is then used to improve the quality of the policies imposed by PCRF.

The policies imposed by call admission control have its effects directly in users that can be affected in different ways. The possibility of doing an handover to another technology that is not congested is one of the solutions that have great advantages. The network provider can give use to all its resources (by using other technologies that it has available and perform handover inter-technology [23]), and keep the connection profitable (since the connection bandwidth is not reduce or dropped). One handover solution can be seen in [33], that uses ANDSF to connect to heterogeneous networks. The solution proposed in [16] uses WiFi network to augment 3G networks.
Other solutions to keep the network running without problems is the use of policies that limit the bandwidth of the users. With the reduction of the throughput available to the users in cases of congestion, the overload of the network will reduce and the network will become less congested. In extreme cases it is even possible to completely remove users from the NodeB to guarantee the end of congestion. The criteria to choose the users to be affected can have different factors into account, such as, the operator options (protect the clients that pay more and their QoS) or choose the heavy users (the main contributors to congestion) in the cell to be the ones to be reduced.

The IpRaft [48] is a solution of PTinovação that is able to implement a Call Admission Control module, since it has in its architecture a PCRF module. This solution is described in detail in the next chapter, since it is going to be used in this Dissertation to implement new policies into the network, based on information delivered by probing and metrics evaluation.

Movik has a solution called REACH [41] (Report, Export, Act, Control and Hetnet) that offers an intelligent RAN control to increase QoE. Report is in charge of obtaining RAN information from multiple sources. It provides information on users, applications, signalling and other sources. Export is the phase that is able to give the information from the report phase to the PCRF using a PCEF module. The third phase is called Act and can control the admissions in the network and also force a reduction of traffic in congested sectors. The next stage is Control, which has the information on the radio access technologies and can control which is the best network to connect. The final phase is HetNet that offers a multi-RAT heterogeneous overlay network to optimize the use of the network, and with the objective to reduce CAPEX to the operator and simultaneously increase the user QoE.

Sandvine [53] has a solution that controls both data plane and control plane using a PCEF and PCRF solutions. It uses a proprietary module, called Policy Traffic Switch (PTS), that uses DPI to identify and measure traffic from clients. The Service Delivery Engine (SDE) that uses a PCRF is in charge of delivering the policies into the network. The final functioning is made using the PTS that deals with the real-time decisions while the SDE manages the network and forces its decisions into the gateways.

2.7.1 Call Admission Control Systems

In this section, it will be described call admission control systems created by different vendors. Their solutions use the architectures modules that were described previously in this Dissertation and also other methods to identify congestion in the network and also infer the QoS of their subscribers.

2.7.1.1 Intelligent Traffic Management

Alcatel-Lucent has a very complete solution to avoid, detect and resolve congestion. The main focuses of this solution is to do an detailed acquisition of information of the network functioning and then impose its policies. This solution is divided into two modules, Alcatel-Lucent 9900 Wireless Network Guardian that has the objective of assemble information in the network and the Alcatel-Lucent 5780 Dynamic Services Controller that implements the policies in the network, the combination of these modules is called Alcatel-Lucent Intelligent Traffic Management.

The Alcatel-Lucent 9900 Wireless Network Guardian [9] provides detection of abnormal network behaviour, performance monitoring and traffic analysis. It monitors all the packets switched in the network to achieved the data rates transmitted and understand how every flow, user and service affect the limited network resources. It also characterizes the most common IP applications used nowadays, providing information on theirs expected impact in the network.

The Alcatel-Lucent 5780 Dynamic Services Controller [8] is a decision engine that constructs rules to monetize the network. This solution uses a PCRF module to create and apply dynamic rules. Its policies are based in combined information from different the network, like, type of UE, radio access network, location of the user, its type of service, credit available or based in parameters like type of service in use or traffic consumptions.

Finally, the Alcatel-Lucent Intelligent Traffic Management [10], the solution that combines both modules described before. This solution enables the service providers to have real time detection of network problems (with the Wireless Network Guardian) and immediately react with the Dynamic Services Controller. With this integration is possible to detect per-subscriber real time events (like defining thresholds to throughputs or data transferred). The detection of heavy users, problems in RAN and low QoS in the network can also be resolved using dynamic policies. The general architecture of this solution is described in figure 2.14.

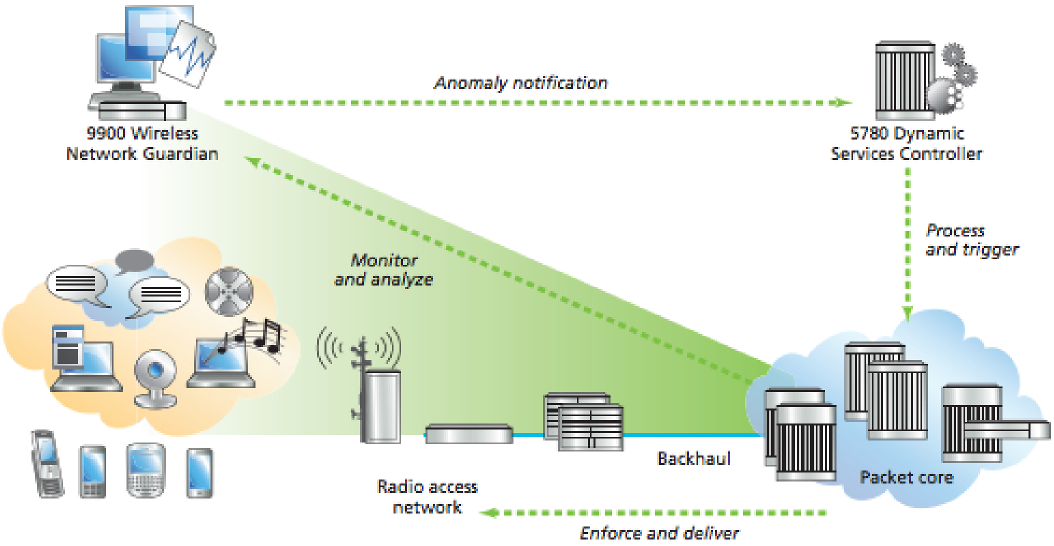

Figure 2.14: Alcatel-Lucent Solution

2.7.1.2 Tekelec

The Tekelec solution for call admission control is based on the detection of the RAN congestion and also the necessity of maintain an ideal QoE for the users [55]. The importance of QoE is crucial to maintain the loyalty of the client, therefore it must be assessed with detailed information on the user behaviour. The factors that can influence the decisions of this solution can be seen in figure 2.15.

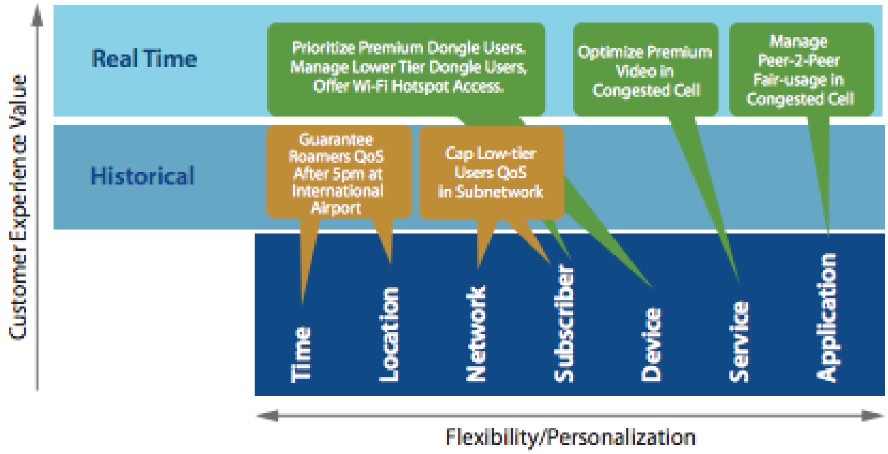

Figure 2.15: Tekelec Management Policies [55]

In this solution, it is used a policy based in three basic steps: analyse, decide and act. With this method, Tekelec proposes a real time feeback system, where it keeps monitoring the network and adjust the resources to every costumer. The evaluation of the network is made using PIC (Tekelec's Performance Intelligence module, described in the architectural figure 2.16) that is capable of performing network probing, assemble network metrics and define RAN alarms to detect congestion to ensure the best quality of service to users. To know more about users, Tekelec uses Tekelec's Subscribed Data Management (SDM) system and with that, it decides how the traffic must be handled in real-time for each costumer necessities. To implement it, Tekelec uses its advanced PCRF module with policies based on user profile combined with RAN aware policies. The final stage comprehends taking actions in the network, and it is made with the concern of prioritizing latency-sensitive traffic (like stream content) and also protect the traffic from important clients (called premium subscribers in Tekelec context); this ensures an overall high quality for services in its network. The implementation/architecture of the Tekelec's solution can be seen in figure 2.16.

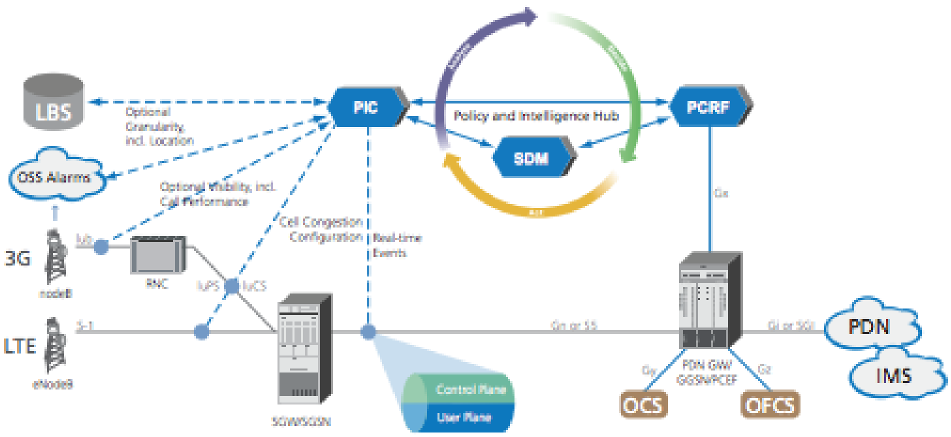

Figure 2.16: Tekelec Architecture Solution [55]

2.7.1.3 Amdoc

The Amdoc solution for call congestion control is based on a Policy Controller with several features capable of configure business rules engine, management of QoS, variation of bandwidth according with traffic all in real-time. This solution is fully compatible with 3GPP standards for PCRF and is very versatile, since it can be used with several networks (HSPA, LTE, WiMAX) as part of EPC. This solution can be integrated with DPIs, access gateways and also content optimization solutions.
This product offers the creation of rules to apply in the network, based on static and dynamic information. These rules can be constructed on a graphical user interface that allows the creation of policies oriented to the user or application.Some of the main features of this product are related with RAN congestion control permitting application sessions per RAN resource. Concerning the user, it can measure its data flows based on time, volume, application or a combination of the enumerated parameters. This solution contemplates a wide range of use case scenarios to manage network performance and capacity intelligently always focused in revenue. The enforcement point for the policies created can go from access gateways, content optimization servers, DPIs and even the subscribed device. These allow the operator to apply its rules across the complete architecture of the network, from the core to the RAN and also the UE. This solution provides granular control over the bandwidth used by the user, and enables modifications in it, which is a way to control congestion by limiting it [12].

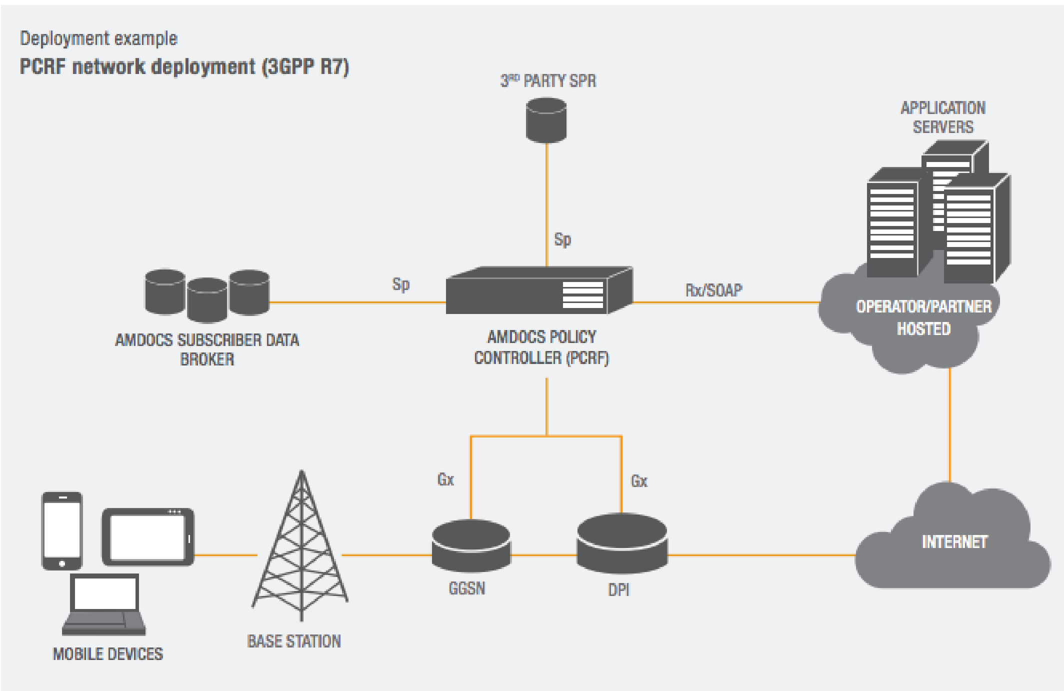

Figure 2.17: Amdoc Architecture Solution [12]

2.7.2 Call Admission Control Algorithms

In [14], it is presented two CAC mechanisms in order to control the admission of VoIP calls, increasing the number of calls answered, taking into account the statistics of previous days for periods of 15 minutes, or based on the current state of the cell. The first CAC mechanism called "BR CAC" is based on reservation of bandwidth for VoIP traffic, given the arrival rate and statistics obtained from previous days. The second mechanism, called "DCAC", unlike the previous one, it does not reserve bandwidth or uses statistical information; instead, it uses the occupancy status of the cell.

In [56], it is presented a generic algorithm that gets information from the QoS requirements of users, the context and status of the various links available from the different access technologies. Based on this information, when the network receives a (new user or handover) request, it chooses the best access technology for that request, taking into account the available resources of each technology and the priority of the request (handover has higher priority). If the access technology of choice does not have sufficient resources, it will try to downgrade the sessions with less priority. If there are available resources, the request is accepted, otherwise it is waiting for a handover, and it rejects the connection if it is a new request.

In [26], it is presented a hybrid and adaptive CAC mechanism (HCAC) for LTE networks. This mechanism has the advantage of using RBs (Resource Blocks). With the use of RBs, the eNB can ensure QoE, and also verify in real-time the occupation and check who is consuming the most (the RBs are assigned according to the type of class of service (QCI), and each class has its own queue). This mechanism gives preference to handovers over new connection request (new users), it prefers GBR traffic over the non-GBR traffic, and considering the GBR traffic it gives preference to the most sensitive traffic. When it does not have enough resources (RBs), it can reduce the RBs associated with non-GBR traffic (starting with those with more RBs) to perform a downgrade.

In [7], it is shown a downgrade policy based on simple CAC mechanism. When a request arises, it checks if the bandwidth is sufficient; in the positive case, it also verifies if that type of traffic has been downgraded/removed and, if this occurs, it does the downgrade/removal. Then, it checks again if there is sufficient bandwidth; if there is not, it is going to downgrade flows, taking into account the type of traffic, to accommodate the request.

In [30] a solution is proposed, that remaps flows of a GBR class into the same class type but non-GBR (modifying the QCI). The process begins when the monitor queue at eNB, detects that the queue of a GBR class exceeds a certain threshold (it is congested) and the queuing of a similar non-GBR class is below the threshold. With this fact, it begins the process of triggering the PCEF through the MME. The PCEF remaps the QCI taking into account the rules in the table and responds to the MME. The MME sends the E-RAB MODIFY REQUEST for the eNB with the new QCI. After that, the queue monitor continues to monitor the queues and signals when it is needed to remap again.

In [36] it is presented a CAC algorithm and resource scheduler. This solution is based on the introduction of a guard interval for transmission, which gives high priority to real-time services, based on the delay deadline. Thus, the algorithm is ready to dynamically adapt to the level of congestion, constantly analyzing the delay. More specifically, the delay of real time traffic is measured constantly, and when it exceeds a defined threshold, the number of RBs allocated to RB traffic is increased, and therefore decremented the RBs for traffic that is not in real-time. In addition, the real-time traffic has a limited in bandwidth, when it is reached, it stops accepting requests.

In [19] an algorithm describing CAC and QoS parameters of the EPS is presented, detailing ARP. The CAC algorithm is based on the information contained in the ARP, and this information can determine the minimum acceptable QoS. With this information, CAC can determine the minimum amount of bandwidth required for services that are already being used, and then decide whether to accept a new request or not. This solution reduces the number of rejected requests, but decreases the QoS of higher priority services.

In [35] it is presented a CAC mechanism that is based on weights assigned to each type of service that vary with network load. To do that, it has an element on each node, called CAC, which monitors and periodically reports the status of the network to a central element of the access network, called "Bandwidth Broker". This central element is responsible for deciding whether or not to accept new orders. Based on the received metrics, it dynamically assigns a weight to each type of service, which involves assigning a width limit hierarchically.

In [6] a CAC algorithm which gives preference to real-time traffic is presented. This mechanism, when there is free bandwidth and it receives a new application request, it is always accepted. When there is not enough bandwidth, it will consider if doing the downgrade of the traffic that is not in real-time is sufficient to accommodate the new request. If it is sufficient, it does the downgrade and accepts the request; otherwise the request it is rejected. Later, when there is available bandwidth, it will upgrade the services to which downgrade had been done.

In [45], it is presented an adaptive CAC algorithm that gives preference to real-time traffic over the traffic that is not in real time. This algorithm assumes that each cell has a fixed number of channels, and that each stream of real-time data occupies a channel. Considering the traffic that is not in real-time, each data stream occupies multiple channels (minimum of one) that varies depending on the occupation of the cell. When a new request comes / handover and the cell is congested, it downgrades the traffic that is not real-time (performing a reduction of allocated channels), to accommodate the request / handover. If the downgrade does not release sufficient resources, the request / handover is rejected. Later, when the cell has free resources, it upgrades the flows that were downgraded before.

In [4] a new idea for the CAC, which involves using congestion prediction is presented. The idea is to store the profile information of the accepted requests, calculate its QoS and store everything in a table, in PCRF. When a request comes to the PCRF, the decision is made based on the policies stored in this table, statistical information and the current state of the network, obtained locally or through monitoring devices in the core nodes.

In [28], it is presented a dynamic CAC algorithm based on QoS, rate of handovers rejected and the rate of new requests rejected. Periodically, the algorithm analyzes the obtained metrics and determines the threshold for handover. If the QoS is above its threshold or rejected handovers rate is above the defined threshold, the handover threshold is reduced by 10%. If the QoS is below its threshold, the rate of rejected handovers is below its threshold, and the rate of new applications rejected is above the defined threshold, the handover threshold is increased by 5 %. Based on this mechanism, the handovers are always accepted, provided they do not exceed the maximum allowed handovers. The new orders are only accepted, since they do not exceed the threshold limit of dynamic handover, and if the real-time traffic does not exceed the maximum limit of requests for this traffic type.

In [42] it is presented a solution for dynamically modifying the QoS, apart from the access network, from the UE, or AF, or network. For this, it is extended the 3GPP PCC architecture with the introduction of a new element called QCMF (QoS Control and Management Function), which has a new interface with the UE, called Rx#, the Rx interface with the AF and the PCRF, and the interface Sp with the SPR. This new component will transform the generic QoS requirements into QoS requirements specific to each access network, making a combination of network capabilities with the requirements of the UE. With this method, it can choose the best access networks for the UE, and can also take into account the state of congestion of it (if a DPI is used to monitor the network).

In [57] a study about monitoring the network in EPS is shown, including the challenges and points in the network where can be done the monitoring through DPI. The study separates the points of observation in SON (Self Organized Networks), control and data. The study refers as monitoring points for SON the interfaces S1-MME, S1-U and X2; to control the S10 and S1-MME interfaces, and for data the SGI, S5/S8 and S1-U interfaces. The study also includes all the information that can be provided by DPI, and an example of the DPI location for a specific project.

The solution proposed in this dissertation will take into account some of the concepts presented in these CAC approaches.

2.8 Conclusions

In this chapter it was described the concepts necessary to understand the solutions that are proposed. The chapter starts with a description of the mobile telecommunications technologies that are going to be studied and used in this Dissertation. After that, it is described in detail the 3GPP network architecture with special focus on the RAN and the PCC components.

Congestion in 3G networks was explained with its main contributors and causes, since it is a problem to be resolved in this Dissertation.

Parameters and methods to evaluate networks were also described in this section; starting with network metrics that can give information on the state of RAN, detailing the events that happen in NodeBs, that are useful to detect congestion. Network Management Systems solutions from different vendors were also shown, with particular interest in Altaia, that takes part in this Dissertation.

Probing is a method used to achieve rich information of the network, like QoS and QoE, that takes part in this proposed solution (Meo Go probing). Therefore, vendors solutions for probing are presented, to acknowledge its features and its performance in network eval-

uation.
Finally, it was shown the Call Admission Control architecture and functioning, specially its integration with methods to achieve network information and the use of PCRF to create dynamic rules. It was shown several vendors products of CAC, since this Dissertation solution integrates and enriches the CAC system of PT Inovação.

Chapter 3

Architecture

3.1 Introduction

This chapter starts with the description of the already existing architecture of PT Inovação platform, the IpRaft. This platform contains a Call Admission Control module, which is going to be improved to support the new functionalities proposed in this Dissertation. The proposed solution describes in detail the new functionalities that are going to be developed, from the RAN congestion detection module to the probe development to acquire the QoS of a service in the network. The integration of the solution in the already existent platform is also detailed. The final section of the chapter is dedicated to show several use cases, which represent the solution that is going to be implemented and what policies it imposes in the network

- Section 3.2: makes an overview of the ip-Raft architecture with a description of its components with special detail on the principal modules that are going to be used in this Dissertation.
- Section 3.3: shows the problem that this Dissertation proposes to solve and how it is going to be solved.
- Section 3.4: describes several use cases that show different problems in the network and the resolutions available for them.

3.2 Overview of IpRaft Architecture

IpRaft is a network Policy solution of PT Inovação that manages the network resources with the objective of giving to the users a better quality of service (QoS) and make the operator resources more cost efficient. IpRaft works like a Policy and Charging Rules Function (PCRF): it can access subscriber databases and other specialized functions and force its policies in real-time directly in the network/user. The IpRaft supports the dynamic creation of rules that makes this solution very versatile and highly flexible, capable of intervene in a wide range of scenarios. In figure 3.1 is possible to observe the high-level overview of IpRaft, the principal elements are the Real Time Signalling Controller, Policy Decision Rules Engine and SPR database and Ipraft Policy Manager [48].

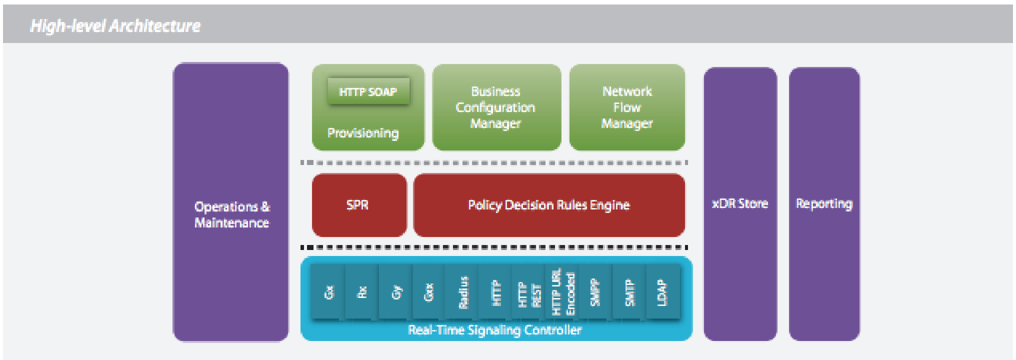

Figure 3.1: IpRaft Architecture [48]

3.2.1 Real Time Signalling Controller

The Real Time Signalling Controller is a signalling gateway that can communicate using different protocols. The main function of this controller is to feed the various modules inside the IpRaft with real-time information coming from different interfaces like Gx, Rx, Gy, Gxx and can use various protocols like Diameter, Radius, HTTP (REST/SOAP) among others.

3.2.2 Subscription Profile Repository

The Subscription Profile Repository (SPR) is a database constructed in MongoDB that has information related with users and their types of subscription. Its information is used to create subscription based policies and rules by PCRF. The fields associated with each element (subscriber) of the database are described in Figure 3.2.

3.2.3 Ipraft Policy Manager

Ipraft Policy Manager (IPM) is a module inside the Policy Decision Rules Engine that is capable of evaluating and determining the rules that should be implemented to the subscribers connection by analysing a wide range of different informations. It can access SPR to get information about users, that is used to choose different policies to be applied. It is fed by the real time signalling controller with network messages, such as Diameter messages that can request the creation, modification or deletion of bearers.

3.2.4 IpRaft Runtime Processor Core

The module IpRaft Runtime Processor Core (IRP-Core) is a sub-component of IPM that is responsible for processing the events that arise in the IPM. The events come from

Figure 3.2: SPR

QDF (Quantum Delivery Framework), and the first module that receives the events is the Protocol Mappers. This module is in charge to evaluate if the network protocol is the expected to a determined service. The middle piece is called Service Criterias, and has the function to analyse if the event must be processed by the service. The final piece is the Service Sessions, that is in charge of processing the event and apply the rules to the event.

The last module is a rule engine called QRE (Quantum Rule Engine). It is a proprietary

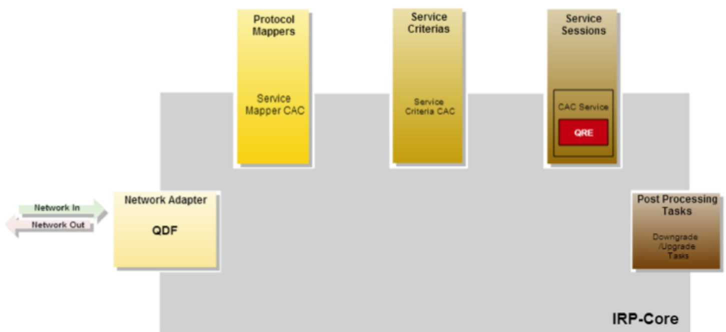

Figure 3.3: IPR-Core

solution developed by PT Inovação that develops and evaluates new rules without the need to recompile its code. This complex actions that this module is able to take are coded in Java (functions like downgrade of the users or drops). These type of actions are usually more lengthy, so after the response to QDF client is done, they are usually processed by the PPT module (Post Processing Tasks).

3.3 Problem and Proposed Solution

To improve the CAC solution, PCRF must detect network congestion in real time, the level of congestion and receive the maximum information available about the users on the congested cell (their ID (MSISDN), throughput, QoS). This information can be divided in two different types, the information related with the network status (the congestion of the cell), and the information related with the subscribers on that cell. The first type of information may come from RAN components; described as metrics in section 2.5 (counters, percentages, etc) that can describe the number of faults, the percentage of occupation, the percentage of availability, among other factors.

PT has a platform that assembles this type of parameters called Altaia, which was presented in section 2.5.1.2. In this solution, a report from Altaia will be evaluated by a module called Metrics Evaluation that will transmit data to the PCRF concerning congestion, so it can force policies into the subscribers to have the network in a stable state.

The proposed solution consists on interpreting the metrics received from the network components and evaluating them to decide if PCRF must be warned about them or not. It is also of great utility to inquire the quality of some specific services provided by PT Inovação to help on the decision process, therefore a probe will be created to enrich the evaluation of congestion on the cell. After this stage, a wider description of the users in the problematic cell must be obtained through probing. This report must contain specific information about the users on the cell, their throughput and the quality of service that they have.

New rulesets must be created, including the metrics and the factors obtained trough probing. The CAC interface must be updated to support the new factors received from this external component (the Metrics Evaluation module), and the CAC decision must apply specific rulesets depending on the information received. An architecture of the proposed solution is depicted in figure 3.4.

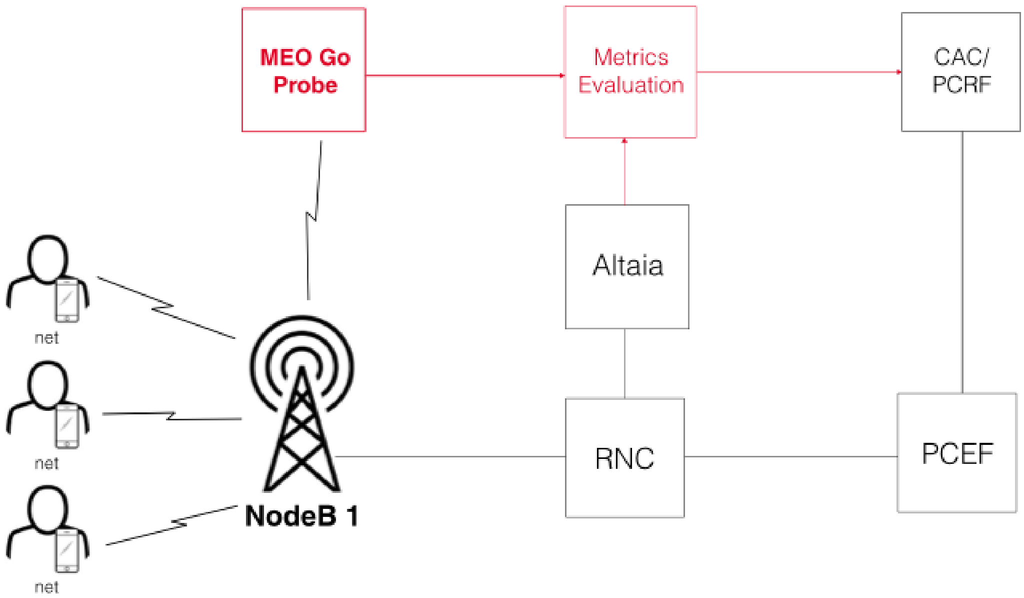

Figure 3.4: Architecture of the proposed solution

3.3.1 Congestion Detection - Metrics Evaluation

Based on the network metrics gathered by the Altaia platform, it is possible to detect congestion on the radio access network. This module will receive a full report with all the metrics about one specific cell. The Metrics Evaluation module is then going to compare the chosen metrics that evaluate congestion and occupancy with predefined thresholds. With an evaluation of the metrics, it will be detected various types of congestion (according with cause-effect) that will be again evaluated to decide the final level of congestion that must be sent to the PCRF. Based on the level of congestion, the PCRF will choose what policies fit the problem detected and then it can decide what type of actions must be sent to PCEF to make the radio access network become non-congested.

3.3.2 Meo Go Probe

The Meo Go Probe has the objective to give information about the service Meo Go on a cell. This probe will be installed on the cells and give QoS experienced on that cell. Factors like latency, bitrate, analyse the network bandwidth availability, errors, buffering state are being evaluated by this probe to give detailed information on the quality of the Meo Go service. Based on this evaluation, depending of its results, PCRF can be notified to act in the network know if the bandwidth of these users can be decreased, and to improve the quality of this service to the users.

3.3.3 Congestion Resolution

The resolution of the network congestion will be done by the PCRF solution previously described. After being notified that congestion is happening in the network, PCRF can give instructions to change the users service plan and consequently reduced the RAN congestion. PCRF actions are selected based on its rules, these rules recognize the level congestion and take different actions depending on how severe the congestion is. The mechanism that defines the admission and control of users is defined in the next sections.

3.3.3.1 Call Admission Control

Call admission control (CAC) is a module integrated in the policy decision rules engine, specifically in IPM, with the particularity of trying to resolve congestion on the radio access network. Figure 3.5 presents an overview of the CAC architecture from its interfaces to its inside modules. The policy implemented by CAC is very complete: it tries to assemble the maximum information about the radio network status, the information about the users on the radio network (typically connect to cells), and the information on subscription plans of these users. With all these informations, CAC is able to be a strong policy manager with the capability of resolving congestion on the RAN.

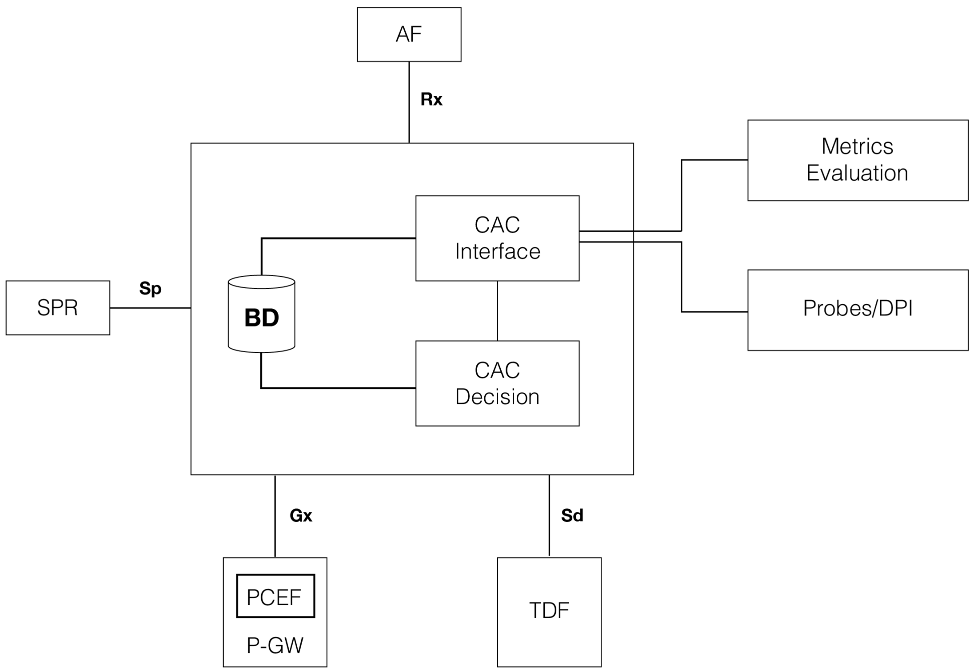

Figure 3.5: CAC

3.3.3.2 Database

The CAC database is used to store information about the cells that are being monitored, and also information about the users on these cells. This data is received from CAC Interface and from other interfaces connected, that after being stored can be used by CAC decision to decide what type of actions the RAN must have to become non-congested. The fields on the database are shown in figure 3.6.

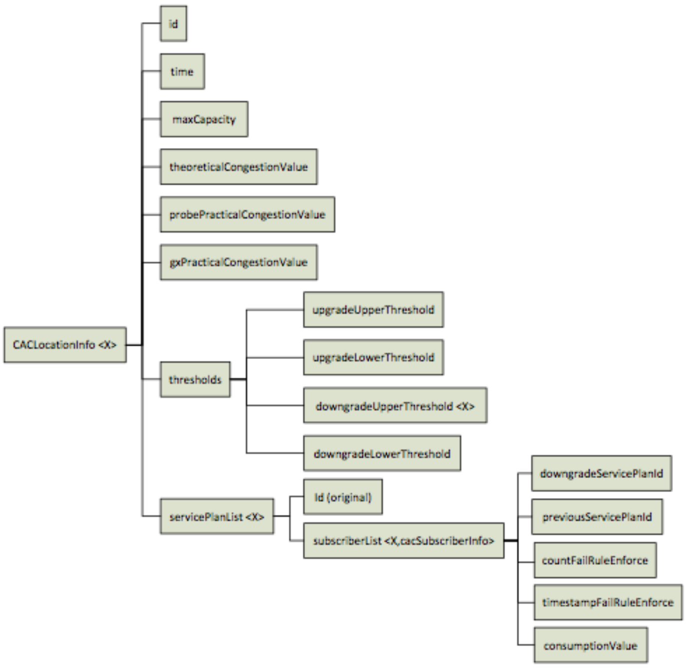

Figure 3.6: Database [37]

3.3.3.3 CAC Interface

The CAC interface is the module that receives the information about the cells congestion that will be stored in the database and posteriorly used by the CAC Decision. The information obtained by this module must be known by it before it arises so it can store it properly in the database.

3.3.3.4 CAC Decision

The CAC Decision is the module that will decide what rules must be applied to different types of users according to the status of the network. It is constructed in QRE (Quantum Rule Engine) but has its complex functions constructed in Java. The QRE section is used to define the rulesets applied to the users, and it is the first approach when congestion occurs; the java section is used to perform complex actions such as, accessing the database

or downgrade/upgrade the users' services.
The downgrades aim to reduce the throughput of the "heavy users" in the cells because they are the main cause of congestion in networks. These happens by changing their service plan to a "lighter service plan" with less impact in the network.The output of this module is sent to PCEF that applies the rules to the users. The CAC decision continues to receive information on the network status; in the case that the network becomes non-congested it can reverse its actions on the users.

3.4 Use Cases

This section will describe several use cases that demonstrate various types of scenarios that can happen in the network. The basic structure of the Use Cases begins with a case of congestion/problems in the network that will be analysed, and in the end, the possible solutions to the problem are showed.

In figure 3.7 it is possible to observe the general functioning of the use cases. At the beginning it is verified if the RAN network is congested or if the services are not having the sufficient bandwidth and quality. If it is true, there are gathered several parameters through probing, that show the users in the cell and all the information needed to achieve their QoS. Then, three solutions are offered as possibilities to resolve the problems: reduction of the users' bandwidth, handover to other technologies available in the area, and finally the drop of services.

The use cases described in the next sections show the ideal functioning of the existing architecture with the modules that were specified in the proposed solution, which implementation is going to be explained in the next chapters. The first two use cases use metrics of a specific vendor, whereas the remaining use cases have the congestion identification done using metrics of the Altaia platform (the first two cases are representative of factors that determine congestion that are not available in Altaia).

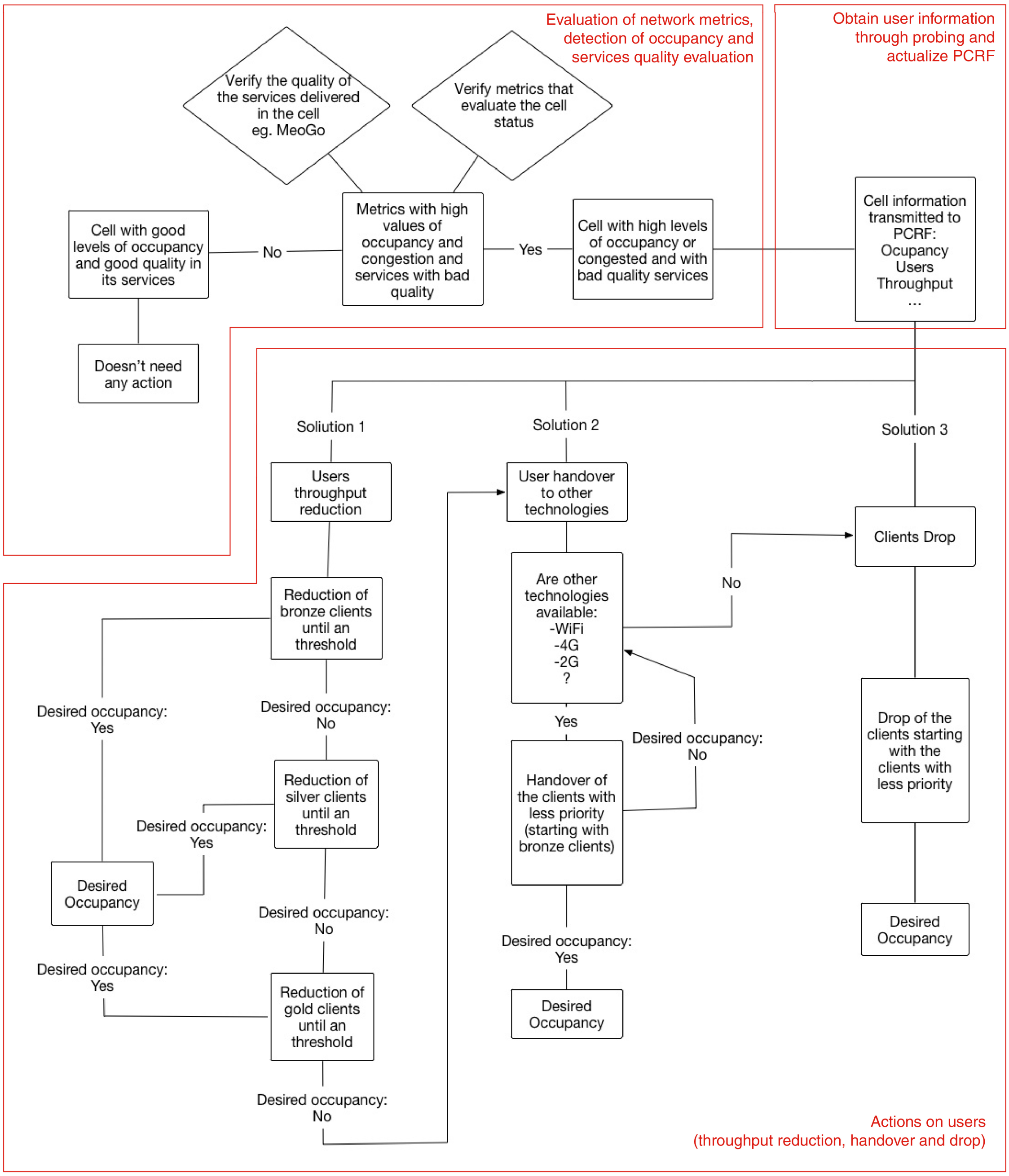

Figure 3.7: Use Cases Flowchart

3.4.1 Use Case 1: Congestion due to Iub bandwidth

On a cell (cell x), several metrics are incremented which shows that there is congestion on the Iub connection. These metrics are the following, presented in table 3.1.

Type	Description	Type of Counter	Value
Number failed RAB, Iub Congestion	Number of failed RAB assignment setup in cell for PS domain, Iub Congestion	Times	20
Average Iub Bandwidth Utilization, UL data	Average utilization rate of the uplink bandwidth at an Iub interface	Percentage	86
Average Iub Bandwidth Utilization, DL data	Average utilization rate of the downlink bandwidth at an Iub interface	Percentage	87

Table 3.1: Use Case 1 Metrics

Evaluating table 3.1, it is possible to see that the occupancy levels of the Iub connection are high and that there are failures in the setup of new connections; therefore, these metrics suggest that the Iub connection is congested.

It is also received information through probing, concerning the general state of the services in the cell and specific information on each user in the cell.
Concerning probing on the cell, ArQos Meo Go probe is used to provide information to evaluate the cell. Since it gives the QoS of Meo Go, it is possible to understand if the cell is having problems by verifying if the QoS is low. The probing results are shown in table 3.2.

Type	Value	Type	Value
Bitrate	300 kbit/s	Perceived Bandwidth	1 Mbit/s
Buffer Audio	13 s	Buffer Video	12 s

Table 3.2: Use Case 1 probe MeoGo

Concerning probing information on each user, it is received the IDs of the users on the congested cell, and also several parameters that allow to verify the experience of the users, presented in table 3.3. In the case that there are many users in the cell, only the ones with higher bandwidth, or the ones with lower requirements in terms of quality, may be shown. The operator will set the policies for all these processes, and here we just include possible examples to show the process provided by the proposed architecture.

MSISDN	Throughput	Transferred Data	Cell ID	RTT	Loss Rate	Application
962321382	2 Mbit/s	200 Mb	cell X	116 ms	0.17%	Browsing
961234567	2 Mbit/s	150 Mb	cell X	132 ms	0.23%	Data transfer
962345678	2 Mbit/s	100 Mb	cell X	123 ms	0.15%	Meo Go
963456789	1 Mbit/s	50 Mb	cell X	117 ms	0.21%	Facebook
964567890	1 Mbit/s	25 Mb	cell X	119 ms	0.20%	Youtube

Table 3.3: Use Case 1 Probing Results

This report is transmitted to PCRF and its data base is updated. With the information available on its database, it is possible to know the client service type, as it is observed in table 3.4.

MSISDN	Maximum Throughput Allowed	Client Type
962321382	2 Mbit/s	Gold
961234567	2 Mbit/s	Gold
962345678	2 Mbit/s	Gold
963456789	1 Mbit/s	Silver
964567890	1 Mbit/s	Silver

Table 3.4: Use Case 1 PCRF

The status is then identified as follows:
87% of NodeB for the Iub DL connection (considering the highest value of occupation);
8 Mbit/s of NodeB are being used;
5 users in cell: 3 Gold + 2 Silver.

Description of the types of service:
Client: Gold;
Service: Net (maximum throughput 2 Mbit/s);

Client: Silver;
Service: Net (maximum throughput 1 Mbit/s);

Client: Bronze;
Service: Net (maximum throughput 0.5 Mbit/s);

With the obtained data it is possible to conclude that all bearers are consuming the maximum data rate available for each one of them, and the occupation of Iub Downlink connection of the NodeB is 87%. The objective is to maintain the occupation under

80%, to be possible to accept new users into the cell (handovers and new connections). Since the element that can be affected is the users' bandwidth, it is made an estimation of the throughput necessary to be released. The calculation is done using the factor that needs to be reduced $87\% - 80\% = 7\%$, and applying it on the bandwidth usage $0.07 \times 8 \approx 0.56Mbit/s$.

With this situation, there are possible solutions that can be approached:

Solution 1

Reduce the data flows of the users in the cell.

With the maximum available bandwidth of the users reduced, it is possible to stop congestion.

One option is to do police/shape the data rates of Silver users, downgrading two Silver clients to the Bronze package, having 0.5 Mbit/s of bandwidth available. The total reduction caused is 1 Mbit/s.

The parameters obtained with probing are used as factors to decide which user must be selected to be downgraded. The users with better QoS within the same service type must be the ones to be downgraded, allowing the ones with lower QoS to increase their experience (equilibrating the QoS of all users). The first criteria to choose the user to downgrade must be the throughput (the client with higher throughput), after it (in case of draw) the Loss Rate (the client with the lower loss rate), and then RTT in case of draw loss rate (the client with lower RTT), and finally transferred Data (the client with higher data transferred). In case of total equality, the users are chosen randomly.

Solution 2

Force the handover of some of the users to different technologies working on that area (e.g. WiFi,2G,...), and check which technologies are not congested on that area to be chosen.

Considering the example of handover to GPRS, it is shown in table 3.5 the metrics that evaluate the handover to this technology.

Type	Description	Type of Counter	Value
Inter-RAT PS Outgoing Handover Attempt number($WCDMA \rightarrow GPRS$) $AttempedHandovers$	Number of attempted incoming PS-domain service handovers to GPRS between systems in a specific cell	Integer	Times
Inter-RAT PS Outgoing Handover Success number($WCDMA \rightarrow GPRS$) $SuccessfulHandovers$	Number of successful incoming PS-domain service handovers to GPRS between systems in a specific cell	Integer	Times

Table 3.5: Use Case 1 Handover Evaluation

Considering the previous information, and the following equation to analyse the percentage of successful previous handovers to GPRS, the GPRS cell is evaluated, to check if it is congested or if it will support the handover.
$S = SuccessPercentage$
$s = SuccessfulHandovers$
$A = AttempedHandovers$

$$S = \frac{s}{A} \times 100$$

Option: By doing the handover of one Silver client to the GPRS technology, the bandwidth used by it will become free, releasing 1 Mbit/s of bandwidth in NodeB.

Solution 3

The third option consists in the removal of users on the cell. In this case, the removal of 1 Silver user of the congested cell will free 1 Mbit/s.

In this case, contrarily to what happens in Solution 1, the criteria to choose the client that must be dropped is related with the worst conditions experienced. The hierarchy shown in solution 1 will be the same to choose the client to be dropped, but now with the inverse value for the factors, starting with throughput (the client with worst throughput), then the Loss Rate (the client with higher loss rate), then the client with the higher RTT and finally the client with the lowest value for transferred data.

The solutions proposed may not be sufficient to achieve the ideal occupancy of the congested cell at the first try, therefore, this can occur as an iterative process. The consecutive reductions will progressively reduce occupancy and stop congestion.

3.4.2 Use Case 2: Congestion due to lack of power

On a cell (cell x), several metrics are incremented which shows that there is congestion on the transmission power. These metrics are the following, presented in table 3.6.

Type	Description	Type of Counter	Value
Number of failed RAB assignment setup in cell for PS domain, Power Resource Congestion	The number of RAB setup failures in specified cell for PS domain service due to transmission power limited	Times	20
Power Load, Average Cell TCP Utilization Ratio	Average utilization rate of the carrier transmission power in a cell	Percentage	88

Table 3.6: Use Case 2 Metrics

Evaluating table 3.6, it is possible to see that the utilization ratio of the used power is high and that there are failures in the setup of new connections, therefore these metrics suggest that the cell is congested.

It is also received information through probing, concerning the general state of the services in the cell and specific information on each user in the cell.
Concerning probing on the cell, ArQos Meo Go probe is used to provide information to evaluate the cell. Since it gives the QoS of Meo Go, it is possible to understand if the cell is having problems by verifying if the QoS is low. The probing results are shown in table 3.7.

Type	Value	Type	Value
Bitrate	400 kbit/s	Perceived Bandwidth	600 kbit/s
Buffer Audio	17 s	Buffer Video	18 s

Table 3.7: Use Case 2 probe MeoGo

Concerning probing information on each user, it is received the IDs of the users on the congested cell, and also several parameters that allow to verify the experience of the users, presented in table 3.8. In the case that there are many users in the cell, only the ones with higher bandwidth, or the ones with lower requirements in terms of quality, may be shown. The operator will set the policies for all these processes, and here we just include possible examples to show the process provided by the proposed architecture.

MSISDN	Throughput	Transferred Data	Cell ID	RTT	Loss Rate	Application
962321382	1.7 Mbit/s	160Mb	cell X	123 ms	0.24%	Gaming
961234567	2 Mbit/s	190Mb	cell X	98 ms	0.17%	Meo Go
962345678	1.8 Mbit/s	180Mb	cell X	117 ms	0.19 %	Skype
963456789	1 Mbit/s	30Mb	cell X	124 ms	0.22 %	Browsing
964567890	0.5 Mbit/s	15Mb	cell X	115 ms	0.14 %	Mail

Table 3.8: Use Case 2 Probing Results

This report is transmitted to PCRF and its data base is updated. With the information available on its database, it is possible to know the client service type, as it is observed in table 3.9.

MSISDN	Maximum Throughput Allowed	Client Type
962321382	2 Mbit/s	Gold
961234567	2 Mbit/s	Gold
962345678	2 Mbit/s	Gold
963456789	1 Mbit/s	Silver
964567890	0.5 Mbit/s	Bronze

Table 3.9: Use Case 2 PCRF

The status is then identified as follows:
88% of NodeB transmission Power is being used;
7 Mbit/s of NodeB are being used;
5 users in cell: 3 Gold + 1 Silver + 1 Bronze.

Type of service:
Client: Gold;
Service: Net (maximum throughput 2 Mbit/s);

Client: Silver;
Service: Net (maximum throughput 1 Mbit/s);

Client: Bronze;
Service: Net (maximum throughput 0.5 Mbit/s);

With the obtained data, it is possible to conclude that the cell is congested due to the lack of transmission power. To be able to receive incoming handovers and new users, the level of this type of congestion must be reduced. A rough estimation of the bandwidth necessary to be released to stop congestion on the cell can be made using the factor that needs to be reduced $88\% - 80\% = 8\%$, and applying it on the bandwidth usage $0.08 \times 7 \approx 0.56 Mbit/s$.

With this situation, there are possible solutions that can be approached:

Solution 1
Reduce the data flows of the users in the cell. By reducing two Silver users to the Bronze package, it is released 1 Mbit/s of bandwidth, decreasing the usage of power in the RAN.

The criteria for the selection of the users that must be downgraded was explained in the previous use case 3.4.1.

Solution 2

Force the handover of some of the users to different technologies working on that area (e.g. WiFi, 2G,...), and check which technologies are not congested on that area to be chosen. The evaluation of this solution was shown in the previous use case scenario solution 2, in section 3.4.1.

Option: Handover of one Bronze and Silver client to other technology.

Solution 3

NodeB removes users to increase the power available in the cell. The selection of the user was explained in the previous use case 3.4.1.

Option: NodeB removes one Bronze and one Silver client.

As in the previous use case 3.4.1, this solution can not guarantee that congestion will be resolved in the first iteration; therefore, in the next report of the network, the solutions can be applied again and progressively decrease congestion.

3.4.3 Use Case 3: Congestion due to lack of Channel Elements (CE)

On a cell (cell x), several metrics are incremented which shows that there is congestion on the channel elements. These metrics are the following, presented in table 3.10.

Type	Description	Type of Counter	Value
CE Cong Setup	Percentage of failed setups due to CE congestion (on the uplink or downlink)	Percentage	12
CE Usage mean DL	Mean usage of channel elements for downlink connections	Percentage	82
CE Usage mean UL	Mean usage of channel elements for uplink connections	Percentage	85

Table 3.10: Use Case 3 Metrics

Evaluating table 3.10, it is possible to see that the occupancy of channel elements is high, and that there are failures in the setup of new connections; therefore these metrics suggest that the cell is congested.

It is also received information through probing, concerning the general state of the services in the cell and specific information on each user in the cell.

Concerning probing on the cell, ArQos Meo Go probe is used to provide information to evaluate the cell. Since it gives the QoS of Meo Go, it is possible to understand if the cell is having problems by verifying if the QoS is low, the probing results are shown in table 3.11.

Type	Value	Type	Value
Bitrate	608 kbit/s	Perceived Bandwidth	650 kbit/s
Buffer Audio	24 s	Buffer Video	22 s

Table 3.11: Use Case 3 probe MeoGo

Concerning probing information on each user, it is received the IDs of the users on the congested cell, and also several parameters that allow to verify the experience of the users, presented in table 3.12.

MSISDN	Throughput	Transferred Data	Cell ID	RTT	Loss Rate	Application
962321382	2 Mbit/s	186Mb	cell X	102 ms	0.19 %	Meo Go
961234567	2 Mbit/s	176Mb	cell X	113 ms	0.20 %	Skype
962345678	0.9 Mbit/s	123Mb	cell X	125 ms	0.23 %	Facebook
962340078	0.8 Mbit/s	100Mb	cell X	123 ms	0.22 %	Youtube
963456789	0.5 Mbit/s	27Mb	cell X	118 ms	0.24 %	Mail
964567890	0.4 Mbit/s	13Mb	cell X	145 ms	0.20 %	Browsing

Table 3.12: Use Case 3 Probing Results

This report is transmitted to PCRF and its data base is updated. With the information available on its database it is possible to know the client service type, as it is observed in table 3.13.

MSISDN	Maximum Throughput Allowed	Client Type
962321382	2 Mbit/s	Gold
961234567	2 Mbit/s	Gold
962345678	1 Mbit/s	Silver
962340078	1 Mbit/s	Silver
963456789	0.5 Mbit/s	Bronze
964567890	0.5 Mbit/s	Bronze

Table 3.13: Use Case 3 PCRF

The status is then identified as follows:
85% of NodeB channel elements are being used;
The banwidth used in the cell is 6.8 Mbit/s
6 users in cell: 2 Gold + 2 Silver + 2 Bronze

Type of service:
Client: Gold;
Service: Net (maximum throughput 2 Mbit/s);

Client: Silver;
Service: Net (maximum throughput 1 Mbit/s);

Client: Bronze;
Service: Net (maximum throughput 0.5 Mbit/s);

With the obtained data,it is possible to conclude that there is congestion due to the lack of channel elements. To resolve this situation, it is necessary that some of the users use less channel elements, therefore it is necessary to reduce the bandwidth utilization on the cell. An estimation on the necessary bandwidth reduction that needs to be released can be obtained doing the percentage occupied minus the desired occupancy, $85\% - 80\% = 5\%$, and applying it to the bandwidth usage $0.05 \times 6.8 \approx 0.3 Mbit/s$.

With this situation, there are possible solutions that can be approached:

Solution 1
Reduce the data flows of the users in the cell. By reducing one Silver user to the Bronze package, it is released 0.5 Mbit/s of bandwidth, decreasing the usage of CE in the RAN. The criteria for the selection of the users that must be downgraded was explained in use case 3.4.1.

Solution 2
Force the handover of some of the users to different technologies working on that area (e.g. WiFi,2G,...), and check which technologies are not congested on that area to be chosen.

Example of handover to GPRS:
Metrics from Altaia that evaluate the state of the cell shown in table 3.14.

Type	Description	Type of Counter
Inter Sys Out HO Succ PS (%)	Percentage of successeful handovers beetween radio access technologies ($WCDMA \rightarrow GPRS/EGPRS$)	Percentage

Table 3.14: Use Case 3 Handover Evaluation

Option: Handover of one Bronze client to another technology.

Solution 3
NodeB removes users to increase the channel elements available in the cell. The selection of the user was explained in the use case 3.4.1.
Option: NodeB removes one Bronze client.

3.4.4 Use Case 4: Congestion due to Code Tree excessive usage

On a cell (cell x), several metrics are incremented, which shows that there is congestion due to the lack of codes. These metrics are the following, presented in table 3.15.

Type	Description	Type of Counter	Value
Code fail (%)	Percentage of fails to setup due to lack of codes NumberOfFails/NumberOfAttempts	Percentage	36
Code Tree Usage Mean (%)	Mean usage of the codes in the cell available in the code tree	Percentage	98

Table 3.15: Use Case 4 Metrics

Evaluating table 3.15, it is possible to see that the utilization ratio of the code tree is very high and it causes failures in the setup of new connections; therefore, these metrics suggest that the cell is congested.

It is also received information through probing, concerning the general state of the services in the cell and specific information on each user in the cell.
Concerning probing on the cell, ArQos Meo Go probe is used to provide information to evaluate the cell. Since it gives the QoS of Meo Go, it is possible to understand if the cell is having problems by verifying if the QoS is low, the probing results are shown in table 3.16.

Type	Value	Type	Value
Bitrate	300 kbit/s	Perceived Bandwidth	330 kbit/s
Buffer Audio	12 s	Buffer Video	10 s

Table 3.16: Use Case 4 probe MeoGo

Concerning probing information on each user, it is received the IDs of the users on the congested cell, and also several parameters that allow to verify the experience of the users, presented in table 3.17.

MSISDN	Throughput	Transferred Data	Cell ID	RTT	Loss Rate	Application
962321382	1.7 Mbit/s	182Mb	cell X	104 ms	0.19 %	Gaming
961234567	2 Mbit/s	196Mb	cell X	108 ms	0.21 %	Skype
962345678	0.9 Mbit/s	123Mb	cell X	121 ms	0.27 %	Youtube
962340078	0.7 Mbit/s	104Mb	cell X	124 ms	0.24 %	Mail
963456789	0.5 Mbit/s	26Mb	cell X	114 ms	0.22 %	Mail
964567890	0.4 Mbit/s	17Mb	cell X	119 ms	0.20 %	Browsing

Table 3.17: Use Case 4 Probing Results

This report is transmitted to PCRF and its database is updated. With the information available on its database, it is possible to know the client service type, as it is observed in table 3.18.

MSISDN	Maximum Throughput Allowed	Client Type
962321382	2 Mbit/s	Gold
961234567	2 Mbit/s	Gold
962345678	1 Mbit/s	Silver
962340078	1 Mbit/s	Silver
963456789	0.5 Mbit/s	Bronze
964567890	0.5 Mbit/s	Bronze

Table 3.18: Use Case 4 PCRF

The status is then identified as follows:
98% of NodeB code tree is being used;
6.2 Mbit/s of the bandwidth;

6 users in cell: 2 Gold + 2 Silver + 2 Bronze

Type of service:
Client: Gold;
Service: Net (maximum throughput 2 Mbit/s);

Client: Silver;
Service: Net (maximum throughput 1 Mbit/s);

Client: Bronze;
Service: Net (maximum throughput 0.5 Mbit/s);

With the obtained data, it is possible to conclude that there is congestion due to the excessive usage of codes. To resolve this situation it is necessary that some of the users use less bandwidth to release some codes. An estimation to the value of bandwidth that must be released can be obtained calculating the percentage of codes used less the desired occupation $98\% - 80\% = 18\%$, and applying it on the bandwidth usage $0.18 \times 6.2 \approx 1.12 Mbit/s$.

With this situation, there are possible solutions that can be approached:

Solution 1
Reduce the data flows of the users in the cell. By reducing one Silver user to the Bronze package and a Gold user to Silver, it is released 1.5 Mbit/s of bandwidth, decreasing the usage of codes by the users. The criteria for the selection of the users that must be downgraded was explained in use case 3.4.1.

Solution 2
Force the handover of some of the users to different technologies working on that area (e.g. WiFi, 2G,...), and check which technologies are not congested on that area to be chosen. This evaluation was made in the previous use case 3.4.3.
Option: Handover of one Bronze client and one Silver client to other technology.

Solution 3
NodeB removes users to increase the codes available in the cell. The selection of the user was explained in the use case in section 3.4.1.
Option: NodeB removes one Bronze and one Silver client.

3.4.5 Use Case 5: Meo Go service problems

On a cell (cell x) , it is received data concerning the service of MeoGo, presented in table 3.19.

Type	Value	Type	Value
Bitrate	300 kbit/s	Perceived Bandwidth	250 kbit/s
Buffer Audio	5 s	Buffer Video	5 s

Table 3.19: Use Case 5 probe MeoGo

The values presented in table 3.19 show that the MeoGo service is having problems, since the buffer values are very low (5 s) and the bitrate is higher than the perceived bandwidth, which means that the service is decreasing the quality (bitrate) or it is in the minimum quality possible. Due to the facts enumerated, it is possible to conclude that the Meo Go service is being provided with a low QoS in the cell.

The ID of the users on the cell is obtained through probing, and also several parameters that indicate the quality of experience of the users, shown in table 3.20.

MSISDN	Throughput	Transferred Data	Cell ID	RTT	Loss Rate	Application
962321382	2 Mbit/s	200Mb	cell X	121 ms	0.15 %	Transfer files
961234567	1.5 Mbit/s	150Mb	cell X	119 ms	0.18 %	MeoGo
962345678	2 Mbit/s	100Mb	cell X	114 ms	0.19 %	Mail
963456789	0.5 Mbit/s	50Mb	cell X	123 ms	0.21 %	MeoGo
964567890	1 Mbit/s	25Mb	cell X	125 ms	0.22 %	Facebook

Table 3.20: Use Case 5 Probe Results

This report is transmitted to PCRF and its data base is updated. With the information available on its database, it is possible to know the client service type, as it is observed in table 3.21.

MSISDN	Maximum Throughput Allowed	Client Type
962321382	2 Mbit/s	Gold
961234567	2 Mbit/s	Gold
962345678	2 Mbit/s	Gold
963456789	1 Mbit/s	Silver
964567890	1 Mbit/s	Silver

Table 3.21: Use Case 5 PCRF

The status is then identified as follows:
7 Mbit/s of NodeB are being used;
5 users in cell: 3 Gold + 2 Silver.

Type of service:
Client: Gold;
Service: Net (maximum throughput 2 Mbit/s);

Client: Silver;
Service: Net (maximum throughput 1 Mbit/s);

Client: Bronze;
Service: Net (maximum throughput 0.5 Mbit/s);

With the obtained data, it is possible to conclude that not all the bearers are consuming the maximum data speed available for each one of them and the QoS of Meo Go is low. The objective is to provide a good Meo Go service to the users in the cell, therefore, the users that are not using Meo Go will be affected.

With this situation, there are possible solutions that can be approached:

Solution 1
Reduce the data flows of the users in the cell that are not using the Meo Go service. By downgrading the users that are not using the service Meo Go, it is possible that the bandwidth released is used by the users that are using it. This allows the network to be managed smartly, since in this case, the streaming application (like Meo Go) requires more bandwidth than the e-mail or facebook. In this case, the user with the ID 964567890 will be downgraded from Silver to Bronze, releasing 0.5 Mbit/s of bandwidth, and the user with the ID 962345678 will have its quality reduced from Gold to Silver, releasing 1 Mbit/s. With this solution, both users using the Meo Go service can increase their data rates and have a better QoS of Meo Go (the users that were reduced, considering the applications that were using, may not notice the reduction of bandwidth).

Solution 2
Force the handover of some of the users to different technologies working on that area (e.g. WiFi, 2G,...), and check which technologies are not congested on that area to be chosen. The evaluation of this solution was shown in the previous use case scenario solution 2 3.4.1.
Option: Handover of the Silver client using the facebook application (1Mbit/s becomes free). With its handover, it is possible to the users using Meo Go to increase their bandwidth and increase their QoS.

Solution 3
NodeB removes the Silver client using facebook of the cell, releasing 1 Mbit/s of bandwidth.

The users using Meo Go may increase the quality of service by increasing the bandwidth available to Meo Go.

3.5 Conclusions

This chapter made an overview of the already existing architecture of IpRaft and its main components, which are very important to this Dissertation, since the implementation of the proposed solution is going to be composed by them. It was also shown the description of the problems that affect the network, the methods that are going to be developed to identify the congestion and acquire information about this congestion, and finally the solutions to the problem. Several use cases were described, presenting different problems that affect networks, and different possible solutions.

Chapter 4

Cell Evaluation and PCRF

4.1 Introduction

This chapter will describe the implementation of part of the entire solution. It starts with a description of how the RAN is going to be evaluated by the Metrics Evaluation Module using the metrics available in the Altaia system. After this stage is described the module that implements the solutions in the network. It starts by describing how the users are simulated in the network, using Seagull software, describing the messages traded to initiate a session, modify it and terminate it. The last section describes the functioning of the CAC solution, the ruleset developed to integrate the Metrics Evaluation module and the applied policies in case of congestion.

- Section 4.2: this section describes the Metrics evaluation module that is in charge of interpret the network parameters, evaluate the status of the network, assign a level of congestion to the cell and transmit it to PCRF;
- Section 4.3: describes the configuration in the existing modules that have to be done so they can support the new parameters;
- Section 4.4: shows how the users are simulated in these modules and the configurations necessary to provide the overall system;
- Section 4.5: shows the functioning of the rulesets implemented in the CAC decision module and how its decisions are transmitted to the users.

4.2 Evaluate Status of the RAN

The first step in this work is the need to detect or anticipate problems that will occur in the radio access network. To determine these factors a program was developed in Java programming language that will obtain the status of a NodeB. This program reads an Excel file, with a report from Altaia which contains a large amount of data concerning one NodeB status that must be processed and analysed.

The fields presented in the report must be compared with another set of fields, previously chosen, the fields of interest (stated in table 4.1) to determine the possible problems of the network. When a field in the report matches a field of interest, the value associated with the factor chosen is compared with a threshold defined for that factor. If this value violates the defined threshold, a value of congestion is attributed (based on the algorithm in figure 4.1). It this case the RAN needs an intervention, therefore the PCRF must be informed to take the right actions to prevent congestion. Before informing PCRF, the program continues to perform a search in the report to evaluate if any other factor of interest was violated. Every time that a field in the received report from Altaia violates a defined threshold, a congestion level is attributed: if it is verified that this case of congestion is worst compared to the previous situations (by comparing the congestion level stored and the level for this situation), this level becomes the worst level of congestion in the cell, therefore is stored. With this process it is guaranteed that the value that will be sent is the one that represents the worst case of congestion.

After checking all the fields referring to one NodeB it is time to check if it is necessary to inform PCRF. If any field was considered to be violating the limits defined, congestion level will be more than 0. The final levels of congestion to be sent to PCRF are described as follows:

- low congestion - level 1;
- medium congestion - level 2;
- high congestion - level 3;

The level 0 is not passed to PCRF since it does not need to act in the network. The simplicity of the parameters transmitted to PCRF is due to the need to process all events as fast as possible. All complexity of detecting and analyse congestion must be implemented in the module described in this section.

After the last evaluation is done, it is generated a XML report with the identification of the cell and the level of congestion it that cell. A HTTP connection with PCRF is created and the report is sent. If no factor was violated and consequently there is no congestion, the program does not report it to PCRF and will just wait for another report of the RAN to come to repeat the process. All this process is described in figure 4.2.

Pseudo Code	
Occupancy Metrics Example: CE usage mean UL(%)	**Failures Metrics** Example: CE Cong setup UL(%)

If(MetricValue > 80%)
Congestion = 1
If(MetricValue > 85%)
Congestion = 2
If(MetricValue > 90%)
Congestion = 3
	If(MetricValue > 5%)
	Congestion = 1
	If(MetricValue > 10%)
	Congestion = 2
	If(MetricValue > 20%)
	Congestion = 3

Figure 4.1: Algorithm to verify congestion with metrics examples

Metric Type	Description
HSDPA Throughput (Kbps)	Throughput in the service HSDPA
HSUPA Throughput (Kbps)	Throughput in the service HSUPA
Code Tree Usage Mean (%)	(Occupancy) Mean usage of the code Tree
CE Usage Mean DL (%)	(Occupancy) Mean usage of CE for downlink
CE Usage Mean UL (%)	(Occupancy) Mean usage of CE for uplink
Code Fails Number (Unid)	(Failures) Number of fails to setup due to lack of channelization codes
Code Fail (%)	(Failures) Percentage of fails to setup due to lack of codes NumberOfFails/NumberOfAttempts
Call CongestionPS (%)	(Failures) Percentage of failed accesses due to lack of resources in the air interface
CE Cong Setup (%)	(Failures) Percentage of failed setups due to CE congestion
Drop RAB PS (%)	(Failures) Percentage of abnormal releases of RAB connection for PS
Call CongestionPS (%)	(Failures) Percentage of failed accesses to RAB PS due to lack of air interface resources
Load Setup Fail Number (Unid)	(Failures) Number of failed setups due to the load of the cell
Load Setup Fail (%)	Percentage of failed setups due to the load of the cell
Inter Sys Out HO Succ PS (%)	Percentage of successeful handovers beetween radio access technologies (WCDMA -> GPRS/EGPRS)
PS Traffic (KB)	Traffic volume in Kb to all PS serivces (R99 + HSPA)

Table 4.1: Metrics

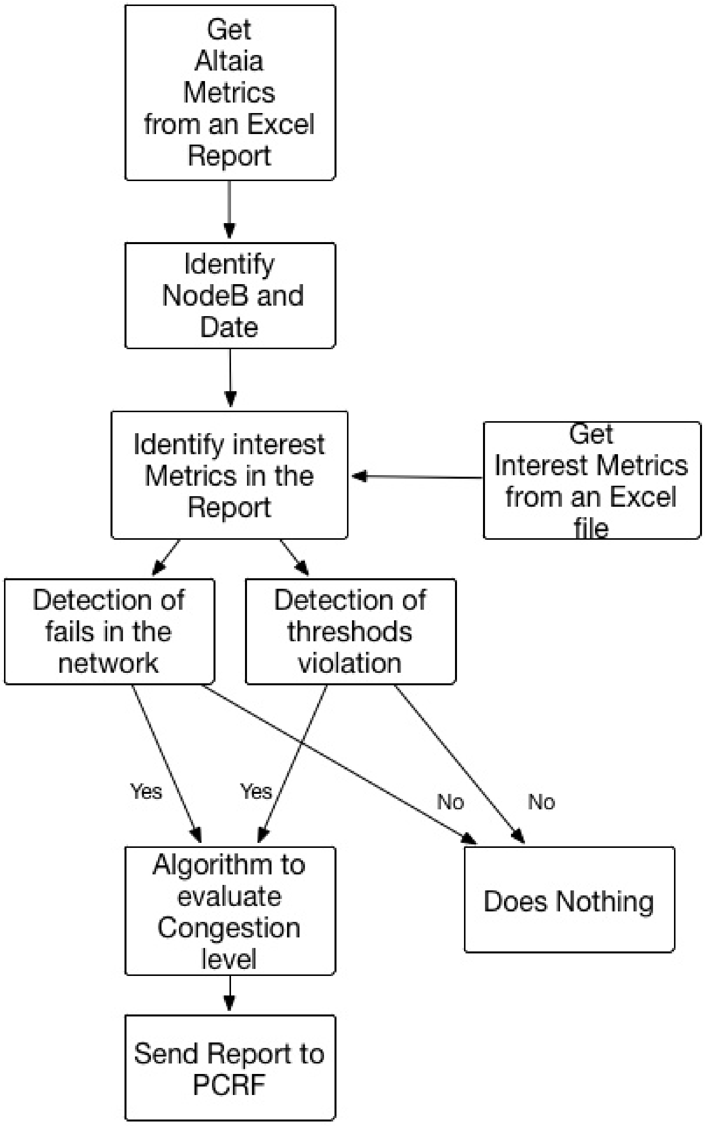

Figure 4.2: Metrics Evaluation Module

4.3 Seagull Configuration

Seagull is an open source software that supports various protocols and generates traffic to simulate users on the network. In this Dissertation Seagull is used to simulate the users that are connected to the cells, and the effects of the actions chosen by PCRF can be checked out on these users. Seagull connects directly on P-GW that will exchange Diameter messages with PCRF, starting a connection between the "user" and PCRF. The messages traded between Seagull and PCRF can be seen in figure 4.3 which shows the process of creating a bearer (when seagull user is connected to the network), the process of modifying a bearer, that affect the user, and finally the process of deleting a bearer when the connection is terminated. The DSGW and DSCF are proprietary solutions of PT Inovação that make the interface between the PCRF and P-GW. The configurations needed at this point are the setup of the messages that are going to be transmitted by the client. It starts with a CCR_INITIAL that has to describe the client (with its MSISDN) according with the information on the CAC database, and also configure the CCR_TERMINATION that is the message sent to terminate the connection.

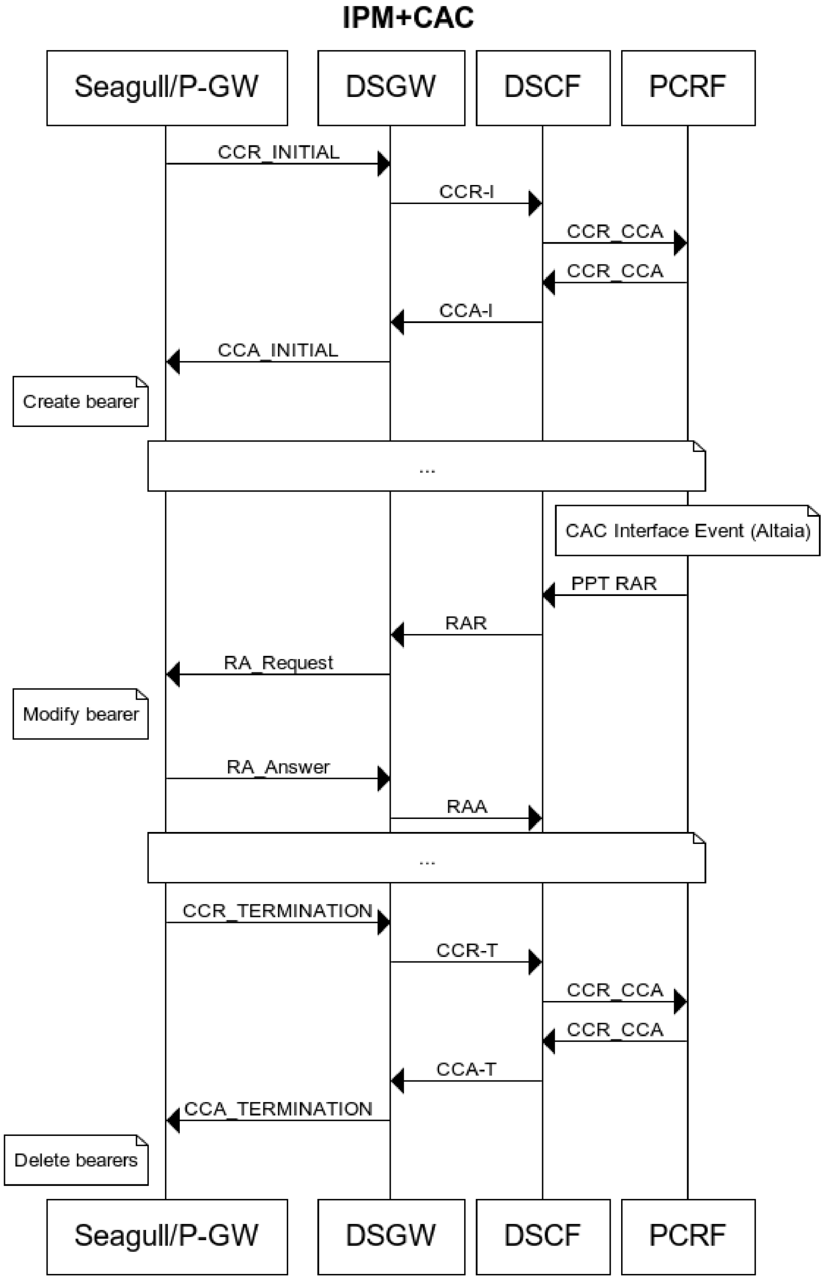

Figure 4.3: Transmitted messages

4.4 Construction of the rulesets

The rulesets are files containing a organized set of rules that will be performed when congestion arises. They are designed in QRE (Quantum Rule Engine) that is a proprietary programming language of PT Inovação. This language also supports code written in Java to perform actions of higher complexity. The main feature of QRE is its flexibility that permits changes in the rulesets without needing to compile them, which makes this module very robust in a real time environment.

The rulesets created aim to have impact on the radio access network and mitigate congestion. The actions that PCRF can force into the network are the following three:

- Reduction of the user throughput;
- Increase of the user throughput;
- Drop the user;

With the use of these three actions, PCRF is able to control and decrease the congestion levels on the RAN. To reduce/increase the throughput of the user, PCRF has to assign a worst/better service plan to this user. For example, if an user is a gold client (the best service plan with the higher throughput) and PCRF needs to free some bandwidth, this client service plan may be reduced to the silver type (a service plan with less bandwidth than the gold).

The rulesets built aim to evaluate the congestion level received from the Metrics evaluation module described in section 4.2, that is on CAC interface. After this step, the parameters received can be used on the construction of the rulesets, in CAC decision.
In figure 4.4 it can be seen the structure of the ruleset, with its description in pseudo code. The ruleset aims to identify the level of congestion, and based on this parameter, decide the severity of the action to be taken. In the ruleset of the figure it is shown that if the cell is highly congested the number of downgrade sessions is higher than if the level of congestion was lower. In this specific case, when the congestion level that arrives in PCRF has the value 1, the number of sessions to downgrade is 5% of all the users in the cell; if the congestion level is 2 the number of sessions to downgrade is 10%, if the congestion level is 3 the sessions to downgrade are 20% of the sessions in the cell.

```
ruleset congestionResolution
rule 1 /parameter('Congestion') is undefined
rule 2 /parameter('Congestion')
rule 3 /CacContext/CacLocationInfo/NumberOfSubscribers
rule 4 /CacContext/CacLocationInfo/
PercentageOfSessionsToDowngrade(rule(3),0.2)
rule 5 /CacContext/CacLocationInfo/
PercentageOfSessionsToDowngrade(rule(3),0.1)
rule 6 /CacContext/CacLocationInfo/
PercentageOfSessionsToDowngrade(rule(3),0.05)
rule 7 /CacContext/StringToDouble(rule(2))
rule 8 /CacContext/sessionsToDowngrade(rule(4))
rule 9 /CacContext/sessionsToDowngrade(rule(5))
rule 10 /CacContext/sessionsToDowngrade(rule(6))
rule 11 if(rule(7)=3) then rule(8) else ""
rule 12 if(rule(7)=2) then rule(9) else rule(11)
rule 13 if(rule(7)=1) then rule(10) else rule(12)
rule 14 if(rule(1) or rule(2)="") then "" else rule(13)
rule 15 /CacContext/CacLocationInfo/
NumberOfPendingActions = 0
rule 16 if(rule(15)) then rule(14) else ""
```

Pseudo Code of the Ruleset (QRE):

If(Congestion == 1)
Number of Sessions to Downgrade = 5% of sessions (users) in the congested cell;

If(Congestion == 2)
Number of Sessions to Downgrade = 10% of sessions (users) in the congested cell;

If(Congestion == 3)
Number of Sessions to Downgrade = 20% of sessions (users) in the congested cell;

Figure 4.4: Ruleset QRE

In the ruleset presented in figure 4.4, it is possible to see references to external code, such as /CacContext/sessionsToDowngrade(1), that reference external functions written in java programming language that will make complex actions, in this example, downgrade the type of session of one user.
From the range of users attached to the problematic cell, some have to be chosen to be affected by PCRF policies. The criteria to choose the users to be affected can rely on different factors:

- Operator Options: the operator can have its own preferences to whom it wants to be affected; criteria like the user service plan (the operator may want to "protect" the users that pay more for their service) can be employed. The operator may also "protect" the QoS of its applications, and therefore choose, users that use other applications to be downgraded.

- QoS: The QoS can be a factor to take into account when it is need to affect users; users with low QoS should not be affected.

- Type of Service: The type of service used by the user is one of the factors that can be used to choose who is going to be downgraded; different services have different characteristics and the decrease of bandwidth affects differently the various types (e.g. the consequences of decreasing the throughput to do the download of an e-mail have less impact that decreasing the throughput available for video streaming);

- Heavy users: The users with higher impact on the network, the main causers of congestion can be selected (based on their throughput) and be the firsts to be affected by PCRF.

Note that these rules are applied to the use cases presented in chapter 3, and will be tested in chapter 6. To this solution the criteria to choose the users that are going to be downgraded is the type of service. The first users to be downgraded are the Bronze, after them the Silver and finally the Gold (with this method the users that pay more tend to have a better service since they are the last ones to be downgraded). In case of draw (more than one user with the some package to be downgraded) the user is chosen randomly.
When it is determined that a downgrade must happen, PCRF sends a RAR (Re-Auth Request) to PCEF to inform that a package should be installed to the user to downgrade. That package is known by the network and characterises the modifications that must be made to downgrade de user. After PCEF executes the process it sends a RAA (Re-Auth Answer) with the answer to PCRF. This process can be seen in figure 4.5 [17].

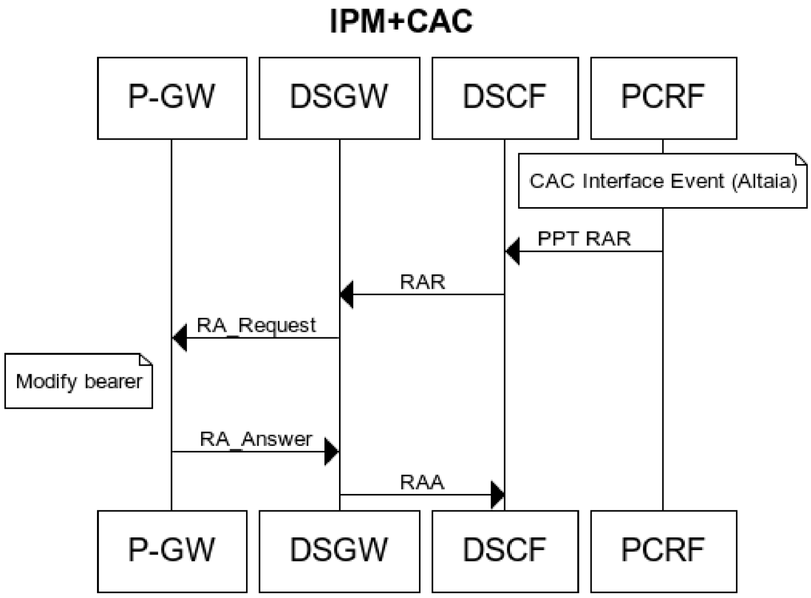

Figure 4.5: Bearer Modification

4.5 Conclusions

This chapter described and implemented the module developed to identify problems in the RAN, mainly focused in congestion. The metrics evaluation module is able to recognize metrics with relevance in congestion detection, and allocate a level of congestion to the RAN that is transmitted to the CAC decision module. The next step in this chapter was the description of the simulation of the users in a cell, that were made using the Seagull software. These Seagull users need to be integrated in the existing architecture of IpRaft to be able to connect to the network and be affected by the network policies implemented. In the last stage, it was described the operation of CAC decision module and the improvement made by using a new ruleset that uses the RAN congestion value. This new ruleset evaluates the level of congestion and, depending on it, it affects a different amount of users that are in the congested cell in a different way. In the implemented rulesets, higher levels of congestion require aggressive solutions to terminate the network congestion. These solutions will be tested in chapter 6.

Chapter 5

Probe Development

5.1 Introduction

Meo Go is an adaptive stream service that provides TV and video on demand to users. It uses a Microsoft solution that provides adaptive streaming called Silverlight. In this section, it is described the implementation of Meo Go Probe that is going to evaluate the service and it will then be used to feed information to the CAC evaluation module. This probe is constructed using several modules of the original streaming service that work together to give an description of the QoS for Meo Go.

This chapter presents an overview of how the probe works and how it is built. A description of the principal modules of the probe is also made: Streaming Module, Download Manager, Buffering Engine and Heuristics Module. The configurations needed to make the probe able to functioning properly were also described. The main requirement for the probe is that it is working in the most similar way possible when compared to the original service, so the QoS results can be accurate.

- Section 5.2: describes the Meo Go service characteristics and its functioning (with detail on the program used to provide video stream);

- Section 5.3: shows the development of the probe and how it is going to play adaptive streams. The principal components of the probe are detailed and an overview of the architecture is provided. The debugging methods and configurations of the probe are also stated.

5.2 Meo Go Service Overview

This section describes the architecture and functioning of the Meo Go service. Meo Go is a service that its main capability is to stream video. To do so, it uses the Microsoft Silverlight to be able to deliver a streaming service. Silverlight is an application framework used to develop internet applications, it is similar to Adobe Flash. It is used as a plug-in for web browsers, and the main feature of this software is its capacity to transmit data with high quality and the possibility to use Full HD content.

Smooth Streaming is a specific service inside the Silverlight environment that is able to deliver adaptive streaming. It is characterized by the capacity to adapt to the network conditions at each moment and deliver a stream with a quality adjusted to the network capacity. This process starts by dividing the video in short fragments, called chunks, that have video content with a duration of around two seconds. The HTTP Web server has the fragments of video encoded using different bitrates that the client application will choose according to its perceived bandwidth. The contiguous chunks are downloaded neatly, stored in the buffer and then played by the order that they were requested. Apart from the different sizes of the chunks (due to the different quality of the video that they contain), they are perfectly synchronized so that they do not overlap each other and the player plays them flawlessly. The main advantages of the smooth streaming media player are:

- It can dynamically adapt to different user equipments and to different bandwidth connections and network conditions, the higher the bandwidth the higher the quality;
- Adaptive streaming uses HTTP caches and proxies, not needing specialized servers at each node;
- Gives a good stream experienced without needing any user interaction;
- Switching from different video qualities is seamlessly to the user;
- It has a fast start-up because the stream always starts with the lower quality and progressively goes to the best quality supported by the network;

Smooth Streaming supports two types of streaming: on-demand and live. On-demand streaming is used to deliver content requested by the user, like movies and TV shows. Live streaming is used to watch channels that are being transmitted live at each moment. The smooth streaming has also the capability of reporting the status of the stream (errors in downloaded chunks, fps,...) and also the status of the network (perceived bandwidth, bitrate,...).

5.3 Probe development

The probe is a module that can evaluate the status of the MeoGo service. It has a complex architecture due to the need to orchestrate several internal modules that must work synchronized.

The implementation is made in C# programming language using Microsoft Visual Studio [39]. Visual Studio is a Microsoft IDE and is used to develop applications for Windows environment. One of the most attractive interest of using Visual Studio is the capability to use the Silverlight platform that has a very rich content concerning internet development and also very useful in video stream development.

This probe has an operation similar to an adaptive stream player, since it has to simulate the behaviour of Meo Go. It has to get pieces of the video to determinate if it is being received correctly, and to be able to give an accurate description of the service.

The adaptive bitrate streaming is a type of streaming that is characterized by the capability of the UE to choose the video quality that it can support considering the state of the network. The video is encoded with different qualities; therefore, the higher the bandwidth the higher the quality of the video received. The stream is delivered through the internet using conventional HTTP servers.

In the beginning of the process, a manifest is downloaded. This manifest is a XML file containing the types of streams available, from the lowest quality stream to the highest quality available for that specific video. It starts to choose the lowest quality available and then, progressively check if it has available bandwidth to upgrade to a better quality stream. A basic description of this process is presented in figure 5.1.

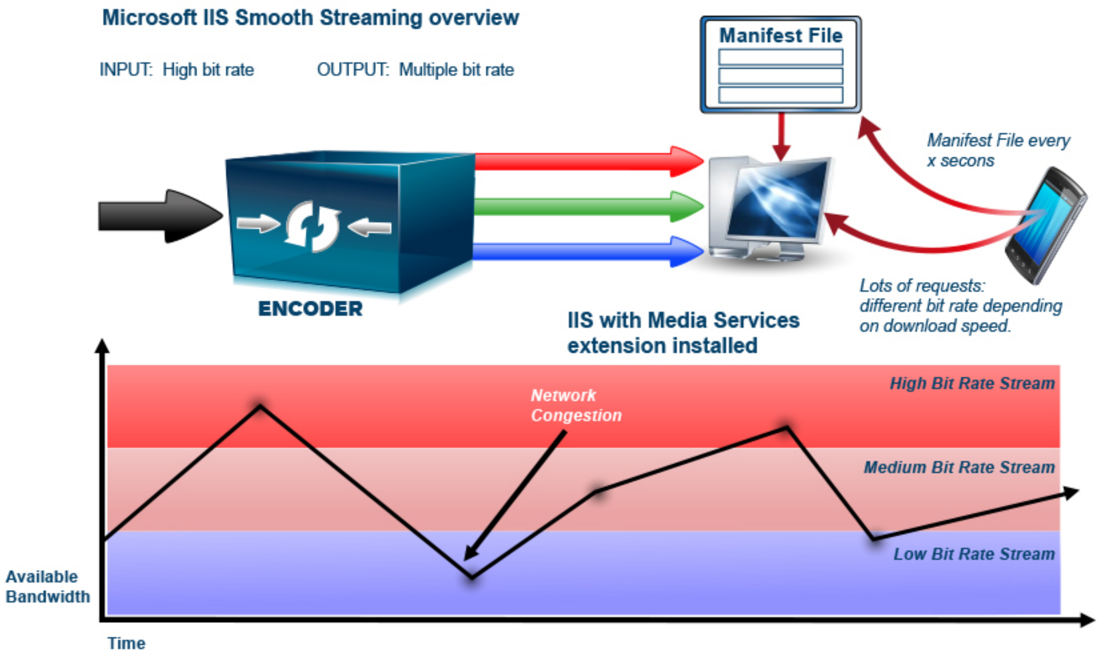

Figure 5.1: Adaptive Streaming [22]

5.3.1 Streaming Source

The Streaming Source is the main class of the probe. It is in charge of controlling all the other modules. It starts by scheduling the download of the manifest. After getting the manifest that describes the chunks and the qualities of the streams available on the server for the stream, it is going to schedule the download of the first chunk. When it is ready to start, it triggers the process of getting new chunks (like if the video was being played); the process of scheduling the download of chunks continues during the entire stream session.

There are other administrative functions in the operation of the probe. It has the function of pausing the player or cancel downloads, or even stop entire modules. Since it is the principal module, it has access to information on all the modules that it controls and also information on the elements controlled by the probe, like manifest information and stream information. The basic functioning of this module and important parts of the probe are depicted in figure 5.2. The Download Manager is the class in charge of downloading the chunks that will be played. Before being played, they are stored in the buffer, using the buffer engine module. The media element is the module that makes the simulated play of the chunks, to simulate the normal function of Silverlight. Periodically, the Network Heuristics module evaluates the functioning of the process and inquires the state of the network; combining both, it determines the bitrate that must be played.

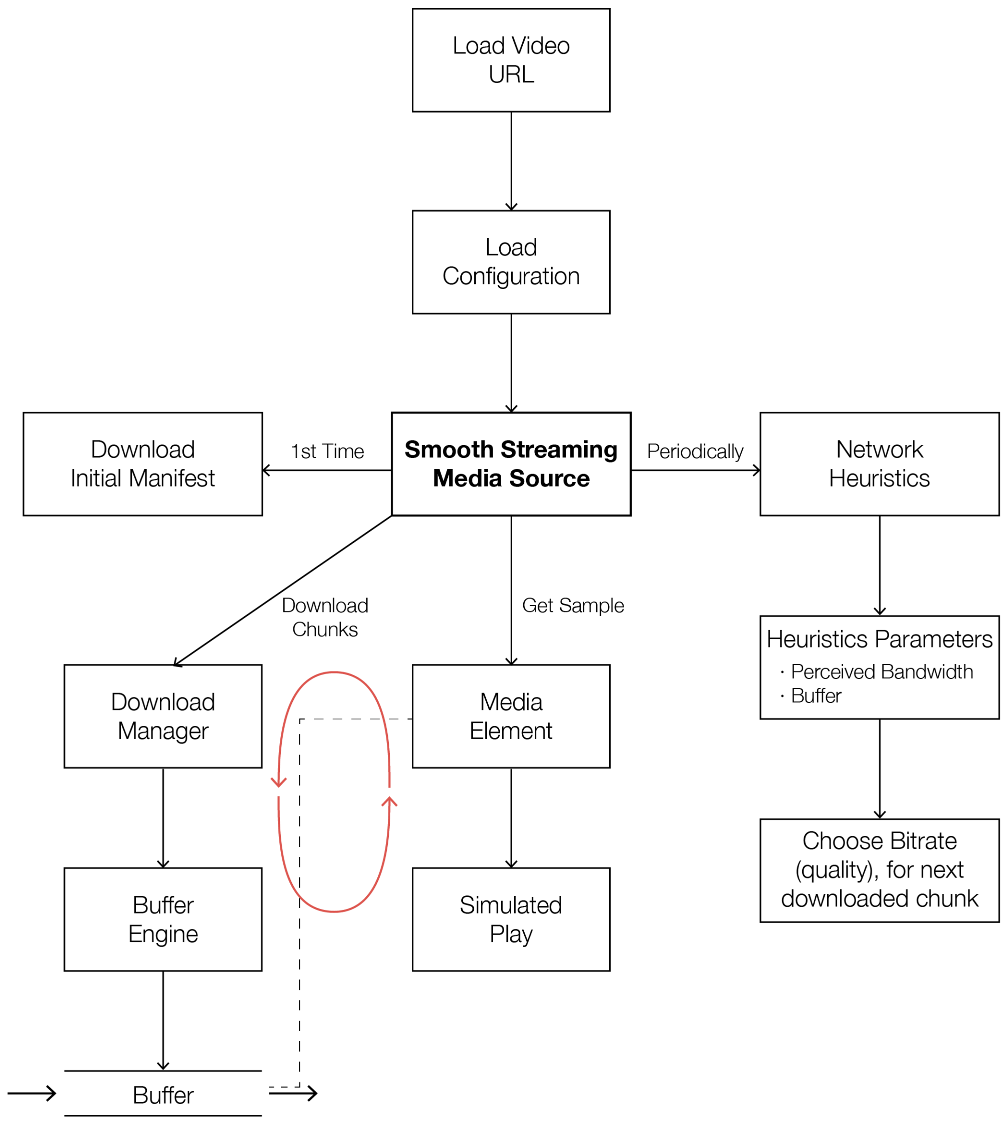

Figure 5.2: Probe Functioning

5.3.2 Download Manager

Download Manager is a class that is capable to download the initial manifest and the various chunks that are needed to assess the QoS of the service. It is constantly being called by the Streaming class to download the next chunks of video that are being required.

This class supports functions to schedule the downloads that are required by the streaming class and several methods directly related with the download. One of the methods is in charge of defining the download source: it receives an URL that is the source of content and the download must be done retrieving information from that URL. It also includes methods like *startDownload* or *cancelDownload* that are directly related with the processing of the downloads. This module can also give reports on the stage of the download. Download Manager contains a method call *UpdateDownloadTrack* that has the function of updating the track from where the download is being made, this method is used specially when the track of the service is switched (the channel is changed).

5.3.3 Media Element

The media element module is in charge to deal with the chunks that were supposed to be played and all the video player configurations/work. Since the objective of the probe is to inquire the state of the network and the quality of service of MeoGo, it does not need to actually play the video. Although it does nott need to play the video the probe needs to receive the chunks and consequently consume them like they are really being played, to the functioning of the probe be similar to the real player. After the chunks have been received they are stored in the buffer, and the samples need to be called to be played (the description of video stream is in figure 5.3). The process of getting samples is described in figure 5.4; after receiving one sample, the probe has to ask for a new one to keep the normal functioning, therefore, when a sample is received the next sample is scheduled to the time when the actual received sample is going to be played.

This module also gives important informations to the other probe modules, like information on global state of the probe, if it is running or is stopped, how long it is be running, the state of the download among other factors. If any error occurs during the use of the chunks, is this class that reports the error to the console terminal.

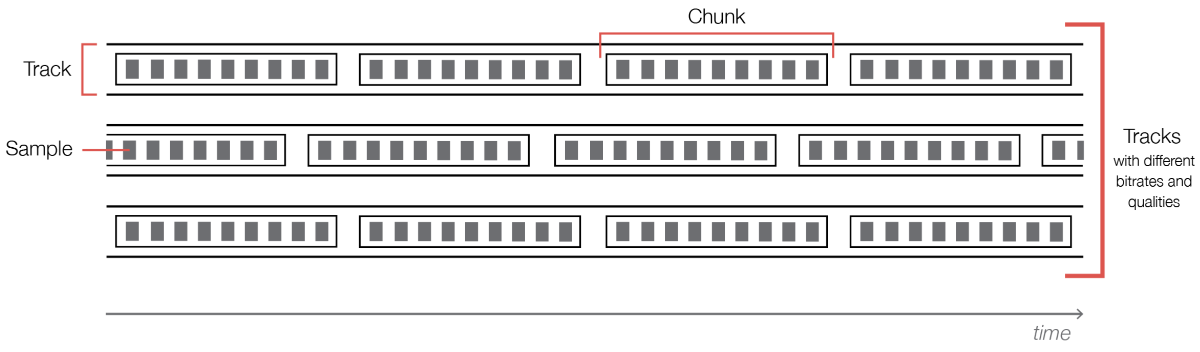

Figure 5.3: Stream elements

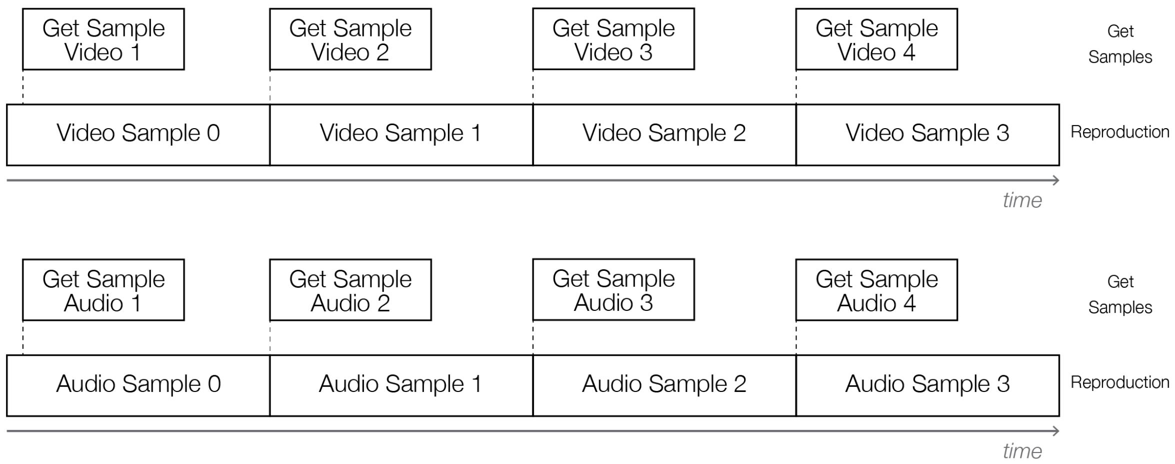

Figure 5.4: Media Element Functioning

5.3.4 Buffer Engine

The buffer engine module controls the buffer usage. The chunks after being downloaded, are stored in the buffer. The buffer has a main rule in the detection of the perceived bandwidth of the network. The perceived bandwidth is obtained by checking the state of the buffer; if the buffer as less elements than a determined threshold, it means that the rate of processing chunks is higher than the rate that the chunks are being received (due to problems in the network such as congestion per example), therefore the quality of the video received must be reduced so the bandwidth capacity can keep up the rate of the chunks being processed.

The Buffer Engine has several methods to manage the buffering operation. It is able to give detailed information on the buffer level in bytes or milliseconds. It has the capability to handle the event of cancelling the download and be able to clean the buffer, so it can support the new elements that are going to be downloaded. Events like download errors or download timeouts are also events that the Buffering Engine module is able to resolve.

5.3.5 Heuristics Module

The Heuristics Module is the section of the probe that is in charge to give the information on the service and the network where the probe is. This can be considered the most important module of the probe, since it is the one that gives the information to the operation of the probe. The other modules on the probe have to simulate the functioning of the stream player, so the Heuristics Module has data to be able to infer the values for the metrics that characterize the service.

The Heuristics Module is composed by several modules with different functions:

- Buffer Heuristics Module: this module evaluates the state of the buffer, specially its occupancy. Based on the occupancy state, the buffer heuristics module suggests a track to be downloaded to the stream: if the buffer is full or near the full state, it suggests an upgrade; if the buffer is in a stable interval, it recommends to keep the track that is being used; if the state of the buffer is empty or bellow the stable state it suggests the reduction of the bitrate.

- Chunk Error Heuristics Module: this module deal with the errors in chuncks, like the chunk requested is not available in cache.

- Latency Heuristics Module: this module evaluates the latency perceived in the probe and, based on it, it recommends a bitrate for the stream.

- Network Heuristics Module: in this module it is evaluated the state of the network. It gives information on the bandwidth perceived by the probe, one of the most important factors to choose one track over another. The larger the bandwidth the larger the bitrate chosen and consequently the quality of the stream.

- Switching Frequency Heuristics Module: this module decides if the stream is able to be changed, based on the need to keep it stable. Basically, it gives a delay to where it is not possible to switch tracks to maintain stability in the quality of the service (do not be always trying to switch on minor improvements of other metrics);

- Track Error Heuristics Module: this module evaluates the errors in tracks, if the bitrate was updated and the new track starts to have errors, it may decide to go back to the previous track and does not allow it to try it again (happens after trying several times);

- Window Size Heuristics Module: depending on the windows size of the screen in the device monitor the window size heuristics module gives a heuristic that determines which is the bitrate that best adequate to the window size. Small window sizes do not require big resolution streams since the difference is not seen by the user.

Each module is in charge of acquiring different output parameters according with each objectives. After that, they are assembled by the Heuristics module. It will produce the following elements describing the video stream elements and the network:

- Bitrate;
- Video Buffer;
- Audio Buffer;
- Perceived Bandwidth;
- Latency;
- Download Errors.

Each module described before, after doing the evaluation of its respective factors, gives the bitrates that it recommends that can be played. In the Heuristics module are then assembled the "reports" of every intern module. Based on them it is achieved the maximum bitrate possible to be played by checking what is the higher bitrate present in all the reports (guaranteeing with this process that every module of the probe recommend the chosen bitrate).

5.3.6 Tracing

The tracing class is used to debug and trace the functioning of the probe. Its main use is to register the logs of the probe in a file, so the developer and user of the probe can follow the main steps and the actual state of the probe. It also has a significant importance during debug of the probe, due to the information/status given about the operation of the probe. It is not shown in figure 5.2 since it is used by every other module of the probe. Every module describes its events by registering them using trace that the trace module writes in a log file.

5.3.7 Configurations

The probe requires some configurations due to its need to simulate the normal functioning of the Meo Go service. To do so, the probe has to be configured with initial parameters like screen size, size of the window on what it is being played the cores available on the probe among other factors. Other parameters are also defined in the beginning of the probe functioning like:

- Audio buffer length in ms;
- Video buffer length in ms;
- Chunk download timeout in s;

- Manifest downlaod timeout in s;
- Bandwidth percentage for upgrade;
- Max Chunk download retries;
- Max audio/video buffer discrepancy between them in ms;

All these parameters must be similar to the ones used in Silverlight, so that the probe has the same behaviour compared to Silverlight.

5.4 Conclusions

In this chapter it was described the functioning of the MeoGo probe and its main components. It starts with the Streaming Source, the main class of the probe that it is the brain of the operation, since it triggers the other modules. The Download Manager is in charge of downloading the chunks that are needed to be played. The necessity of simulating the playback conditions were also described in the Media Element section and its operating mode was detailed. The Buffer module of the probe was also described, and its great importance in the available bandwidth detection was also stated. Finally concerning the main modules, it was described the heuristics module, the module that gives various informations on the state of stream. The tools needed to debug during the construction of the probe were also presented here. A subsection regarding the configurations of the probe, like buffer times or fails oriented parameters were also described in this section.

Chapter 6

Results

6.1 Introduction

This chapter presents the tests and discusses the results of the proposed approaches: from the probing to the metrics evaluation and the results of call admission control. The developed Meo Go probe has been tested in different scenarios and network conditions. The Metrics Evaluation module is tested, to achieve the status of the RAN using real past events on a network. The results of this evaluation show the level of congestion on the network. In the final stage, all the modules are tested together to acquire a congestion level, and send it to PCRF, and check if it enforces the right policies into the network.

- Section 6.2: shows the results of the Meo Go probe tested in different network environments;
- Section 6.3: presents the results of the evaluation of the network, and describes the results obtained after the PCRF takes action to the subscribers simulated with Seagull.
- Section 6.4: shows the evaluation done by the Metrics Evaluation module, and describes the actions taken by the PCRF to the subscribers simulated with virtual machines.

6.2 Probing results

This section shows the results obtained with the probe developed in different test scenarios. The scenarios compare the functioning of the probe with the functioning of the Silverlight program. The various methods to test these elements require a variation of the network bandwidth that is performed using a program called Network Emulator Client, that is capable of defining the bandwidth available. The scenarios show different contents being evaluated by the probe compared to being played on the Silverlight program (the program that provides the Meo Go service).

6.2.1 Video on Demand - Mix video

In this section, it is shown the performance for both Silverlight and probe program running the Mix video. A screenshot of the video can be seen in figure 6.1.

Figure 6.1: Mix Video screenshot

A resume of the manifest of this video is shown in the next frame:

QualityLevel Index=”0” Bitrate=”4000000” Width=”1280” Height=”720”
QualityLevel Index=”1” Bitrate=”3400000” Width=”1280” Height=”720”
QualityLevel Index=”2” Bitrate=”2800000” Width=”1056” Height=”592”
QualityLevel Index=”3” Bitrate=”2200000” Width=”1056” Height=”592”
QualityLevel Index=”4” Bitrate=”1800000” Width=”848” Height=”480”
QualityLevel Index=”5” Bitrate=”1400000” Width=”848” Height=”480”
QualityLevel Index=”6” Bitrate=”1200000” Width=”640” Height=”352”
QualityLevel Index=”7” Bitrate=”1000000” Width=”640” Height=”352”
QualityLevel Index=”8” Bitrate=”800000” Width=”424” Height=”240”
QualityLevel Index=”9” Bitrate=”600000” Width=”424” Height=”240”
QualityLevel Index=”10” Bitrate=”450000” Width=”320” Height=”176”
QualityLevel Index=”11” Bitrate=”300000” Width=”320” Height=”176”

It is possible to see the bitrates (the unit is bit/s) available for the video and the Width and Height for those bitrates. With the variation of the bandwidth available to Silverlight and the probe, it is expected that both adapt to the network conditions and choose the adequate bitrate to stream.

6.2.1.1 Mix video - Downlink 512 kbit/s and Uplink 128 kbit/s

In this test scenario, the probe and the original Silverlight program were included in the same network (Instituto de Telecomunicações network) with an uplink bandwidth of

128 kbit/s and a downlink bandwidth of 512kbit/s. It was requested to achieve the quality of service available in the network. The results obtained for the perceived bandwidth and bitrate can be seen in the next figures.

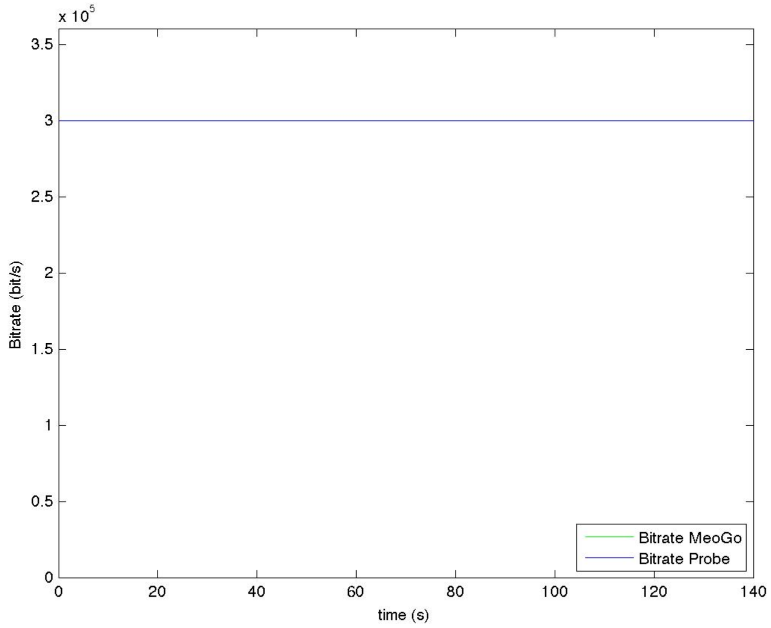

Figure 6.2: Bitrate selected for Mix Video with 512 kbit/s for downlink

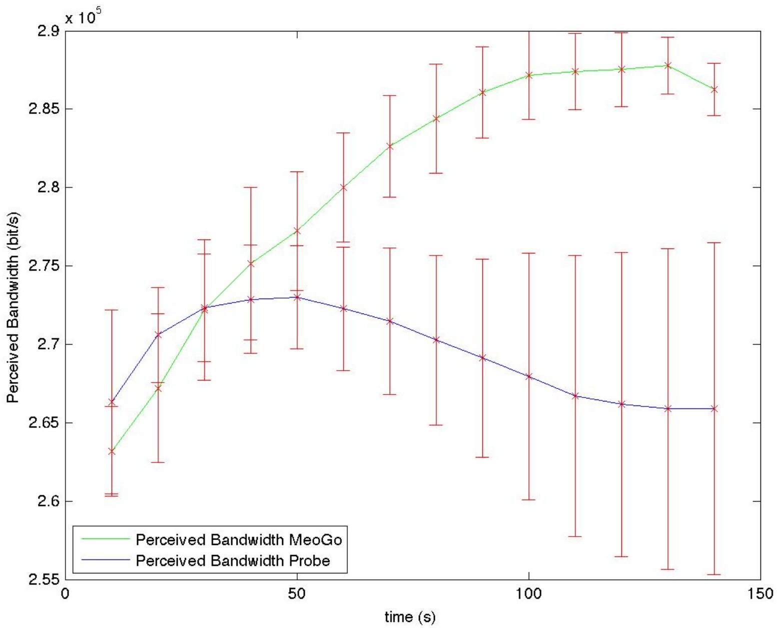

Figure 6.3: Perceived Bandwidth for Mix Video with 512 kbit/s for downlink

In figure 6.2, it is possible to see the chosen bitrate by both probe and Silverlight. It is only possible to see one line because the lines are overlapped. The result of both probe and Silverlight process for bitrate selection determines that 300 kbit/s is the appropriate bitrate to the stream. This bitrate is the lowest bitrate available for the stream, representing the lowest quality available.

The perceived bandwidth by both processes can be seen in figure 6.3. This shows why the bitrate chosen is the lowest, because the bandwidth perceived is approximate to the value of the bitrate. Comparing the results of probe and Meo Go, they are similar, having a discrepancy of less than 25 kbit/s. Compared to the total bandwidth available (512 kbit/s) the bitrate is lower than the vailable bandwidth: this is normal, since the bandwidth has fluctuations, and also the application and TCP/IP cause overhead in the process.

6.2.1.2 Mix video - Downlink 1024 kbit/s and Uplink 256 kbit/s

Similarly to the previous test scenario, the probe and the original Silverlight program were put in the same network but now with an uplink bandwidth of 256 kbit/s and a downlink bandwidth of 1024 kbit/s. It was requested to achieve the quality of service available in the network.

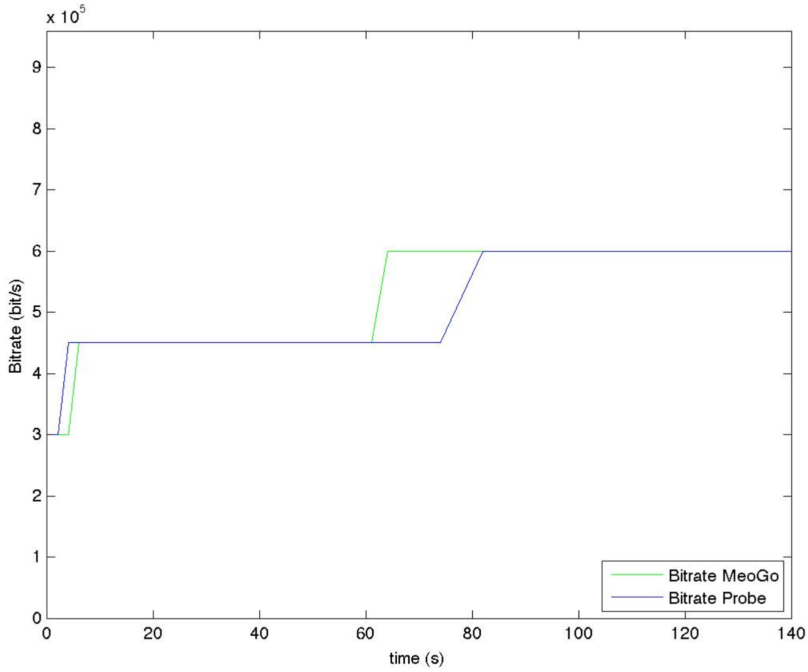

Figure 6.4: Bitrate for Mix Video with 1024 kbit/s for downlink

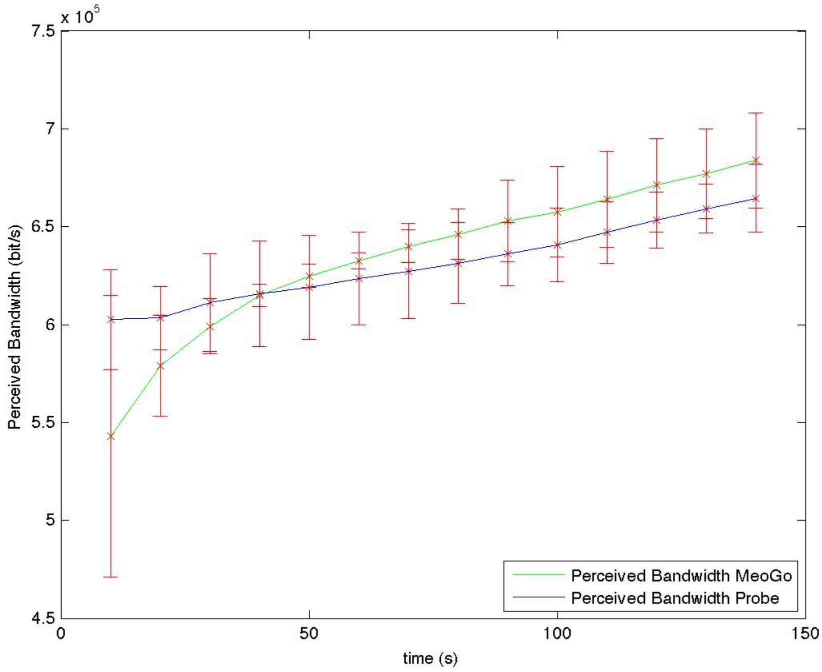

Figure 6.5: Perceived Bandwidth for Mix Video with 1024 kbit/s for downlink

In figure 6.4, it is possible to see the chosen bitrate by both probe and Silverlight. It is possible to see that both processes start by choosing the lower bitrate and it is quickly upgraded to the next quality level. The recognizance of the possibility to upgrade to the 600 kbit/s bitrate takes approximately 60 seconds, and it is accompanied by the increase of the perceived bandwidth.

The perceived bandwidth is shown in figure 6.3 for both modules. The results are very close and present the same behaviour. Compared to the total bandwidth available (1024 kbit/s), the bitrate values are again lower, due to the same reasons as before.

6.2.1.3 Mix video - Downlink 10000 kbit/s and Uplink 5000 kbit/s

In this test scenario, the probe and the original Silverlight program were put in the same network with an uplink bandwidth of 5000 kbit/s and a downlink bandwidth of 10000 kbit/s.

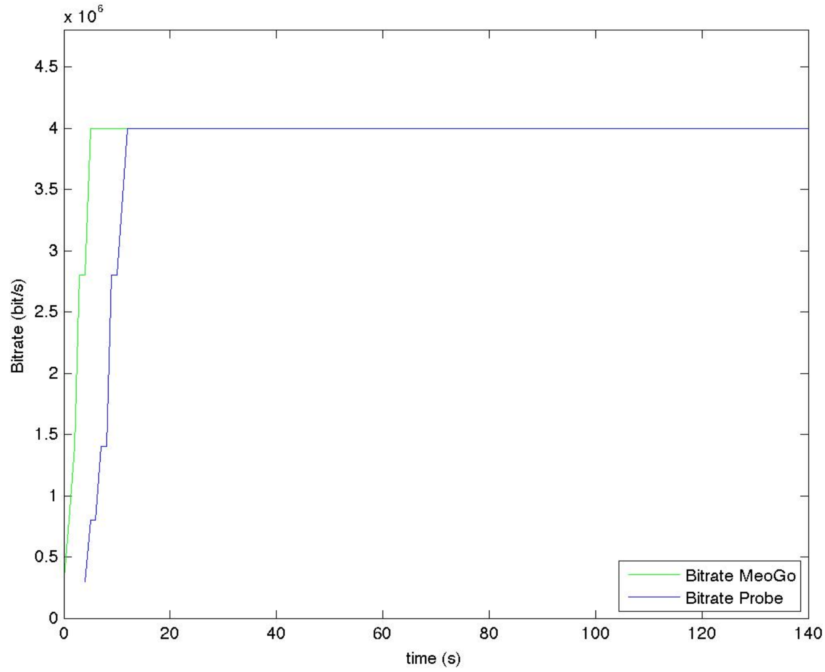

Figure 6.6: Bitrate for Mix Video with 10000 kbit/s for downlink

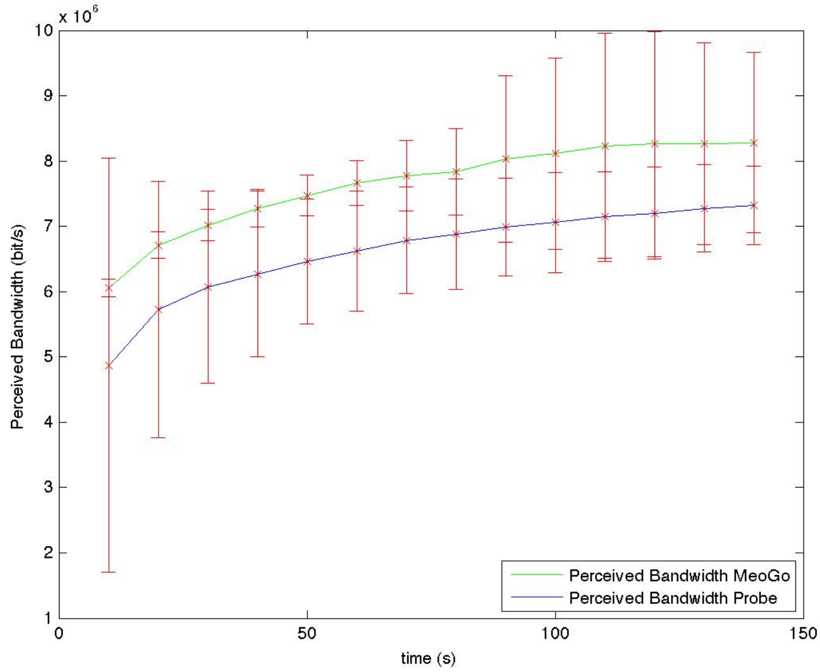

Figure 6.7: Perceived Bandwidth for Mix Video with 10000 kbit/s for downlink

In figure 6.6, it is possible to see the chosen bitrate by both probe and Silverlight. It is possible to remark that both processes choose quickly the best bitrate available, 4 Mbit/s. The upgrade to the best bitrate is made under 10 seconds of time elapsed.

The cause is the perceived bandwidth that is shown in figure 6.3. The results show that from the beginning, probe and Silverlight always recognize a wide bandwidth, larger than the 4 Mbit/s required to the best bitrate, allowing the big increase of bitrate. Comparing the results from both modules in figure 6.7, they are very similar and the mean values are inside each other confidence intervals.

6.2.2 Video on Demand - Big Bunny video

In this section, it is shown the performance for both Silverlight and probe program running the Big Bunny video. A screenshot of the video can be seen in figure 6.8.

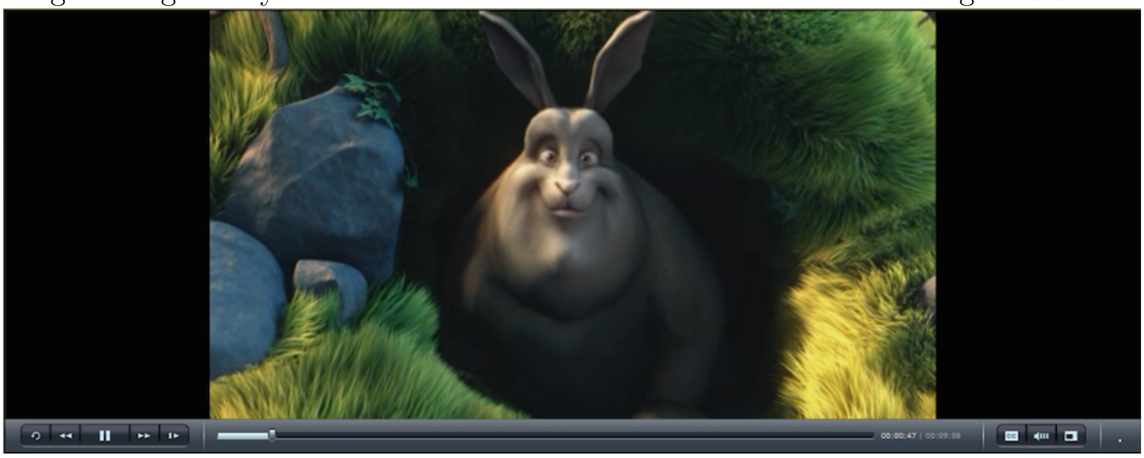

Figure 6.8: Big Bunny video screenshot

A resume of the manifest of this video is shown in the next frame:

QualityLevel Bitrate="2436000" Width="1280" Height="720"
QualityLevel Bitrate="1636000" Width="960" Height="544"
QualityLevel Bitrate="1233000"Width="848" Height="480"
QualityLevel Bitrate="866000" Width="624" Height="352"
QualityLevel Bitrate="608000" Width="480" Height="272"
QualityLevel Bitrate="427000" Width="424" Height="240"
QualityLevel Bitrate="300000" Width="320" Height="176"

It is possible to see the bitrates (the unit is bit/s) available for the video and the Width and Height for those bitrates. With the variation of the bandwidth available to Silverlight and probe, it is expected that both adapt to the network conditions and choose the adequate bitrate to stream. Since the qualities available for this video are different than the qualities available for the Mix video, the expected results are different compared to the previous section.

6.2.2.1 Big Bunny video - Downlink 512 kbit/s and Uplink 128 kbit/s

Similarly to what happened in the scenarios with the Mix video, both probe and Silverlight were now submitted to the same test conditions but with the Big Bunny video. The uplink bandwidth available is 128 kbit/s and the downlink bandwidth is 512kbit/s. It was requested to achieve the quality of service available in the network. The results obtained for the perceived bandwidth and bitrate can be seen in the next figures.

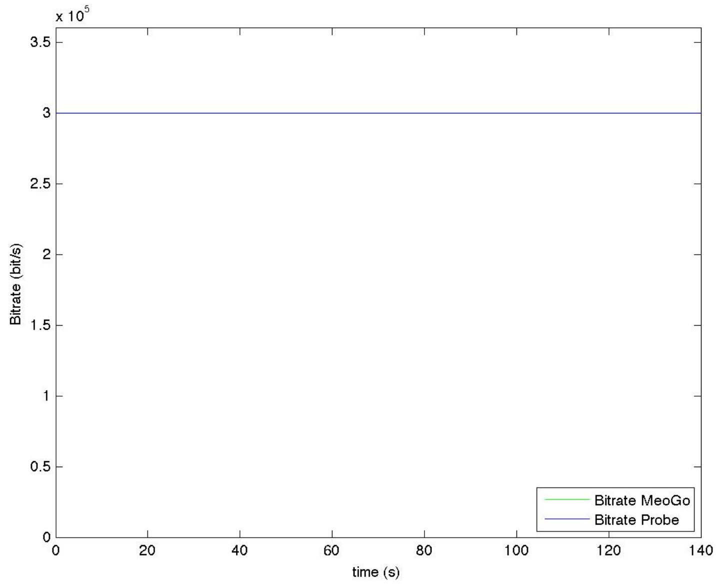

Figure 6.9: Bitrate for Big Bunny Video with 512 kbit/s for downlink

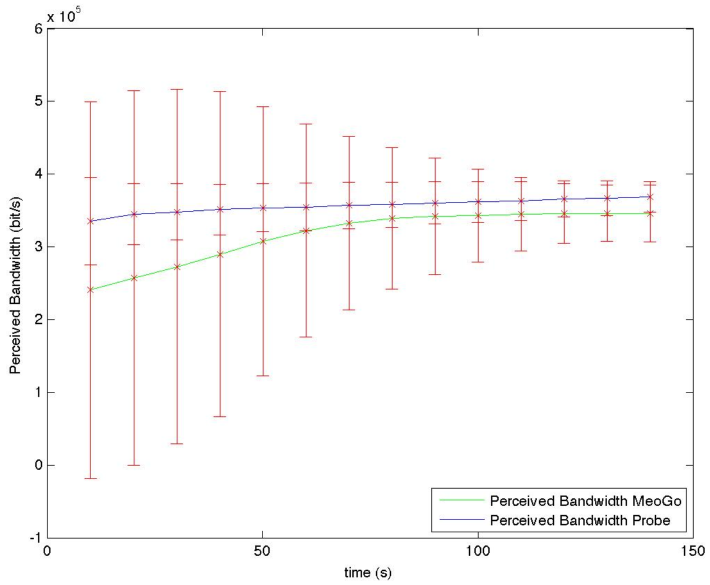

Figure 6.10: Perceived Bandwidth for Big Bunny Video with 512 kbit/s for downlink

The chosen bitrate for the stream is shown in figure 6.9. It is possible to see that both processes (the lines are overlapped) take the minimum bitrate, 300 kbit/s. This is caused by the perceived bandwidth that is shown in figure 6.10. The results show a perceived bandwidth near the value of 300 kbit/s, therefore it is reasonable to choose it for the bitrate. Comparing the results from probe and Silverlight, it is possible to see that they are very close and tend to approximate with time.

6.2.2.2 Big Bunny video - Downlink 1024 kbit/s and Uplink 256 kbit/s

Equivalently to the scenario shown before, both probe and Silverlight were put in the same network conditions with an uplink bandwidth of 256 kbit/s and a downlink bandwidth of 1024 kbit/s. It was requested to achieve the quality of service available in the network. The results obtained for the perceived bandwidth and bitrate can be seen in the next figures.

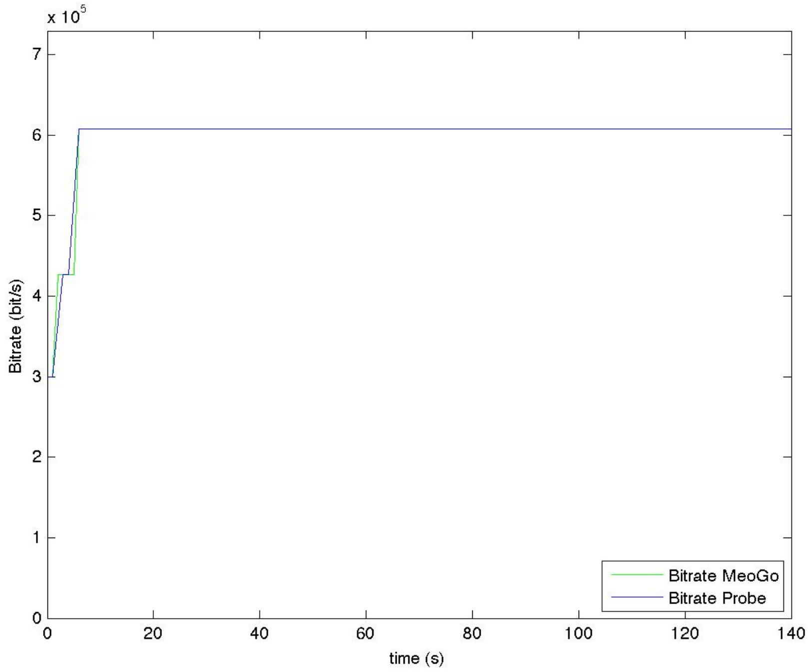

Figure 6.11: Bitrate for Big Bunny Video with 1024 kbit/s for downlink

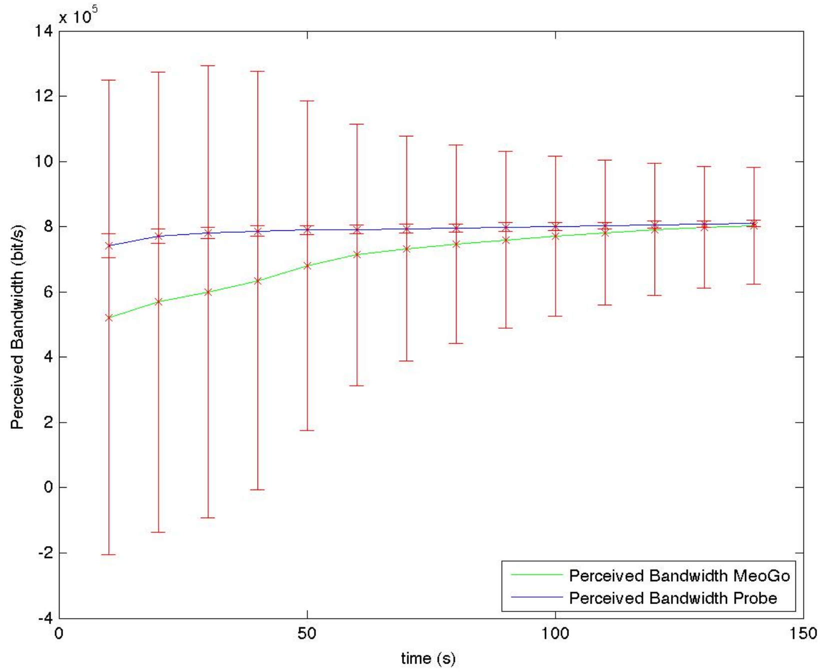

Figure 6.12: Perceived Bandwidth for Big Bunny Video with 1024 kbit/s for downlink

In figure 6.11, it is possible to see the chosen bitrate by probe and Silverlight. It is possible to see that both processes quickly adapt their bitrates according with the perceived bandwidth.

The perceived bandwidth is shown in figure 6.12 and it shows that the values stay almost all the time between 600 kbit/s and 800 kbit/s, justifying the adoption of the 608 kbit/s bitrate. Relatively to the probe and Silverlight evaluation of bandwidth, the results are very similar becoming even closer with time.

6.2.2.3 Big Bunny video - Downlink 10000 kbit/s and Uplink 5000 kbit/s

In this test scenario, the probe and the original Silverlight program were put in a network with an uplink bandwidth of 5000 kbit/s and a downlink bandwidth of 10000 kbit/s.

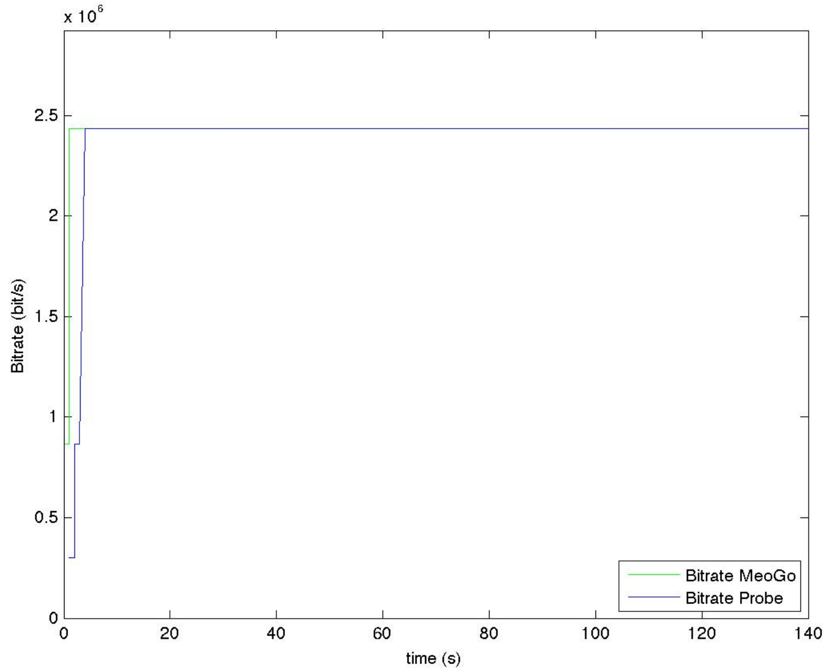

Figure 6.13: Bitrate for Big Bunny Video with 10000 kbit/s for downlink

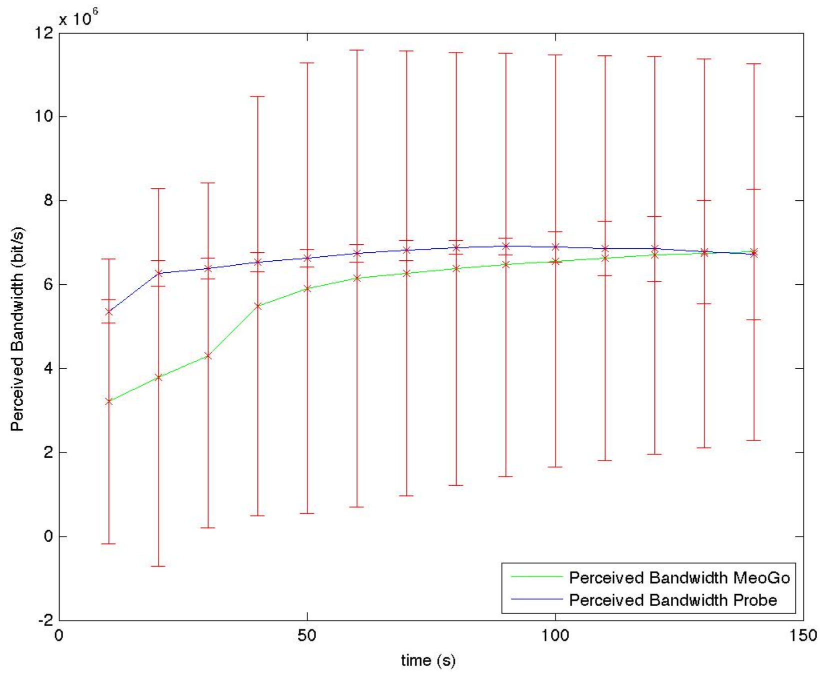

Figure 6.14: Perceived Bandwidth for Big Bunny Video with 10000 kbit/s for downlink

The chosen bitrate for the stream is shown in figure 6.13. In this case, both modules assume the higher bitrate available in short period of time. This immediate increase results in the bandwidth perceived by both modules. Shown in figure 6.14, the perceived bandwidth has values rounding 6 Mbit/s, which allow the bitrate to be the maximum available (since it only needs 2.436 Mbit/s to play the higher quality). Similarly to what happened before, the results of both modules are very close to each other and tend to approximate with time.

6.2.3 Streaming channel

In this section, it is shown the performance for both Silverlight and probe program running a streaming channel. A screenshot of the video can be seen in figure 6.15.

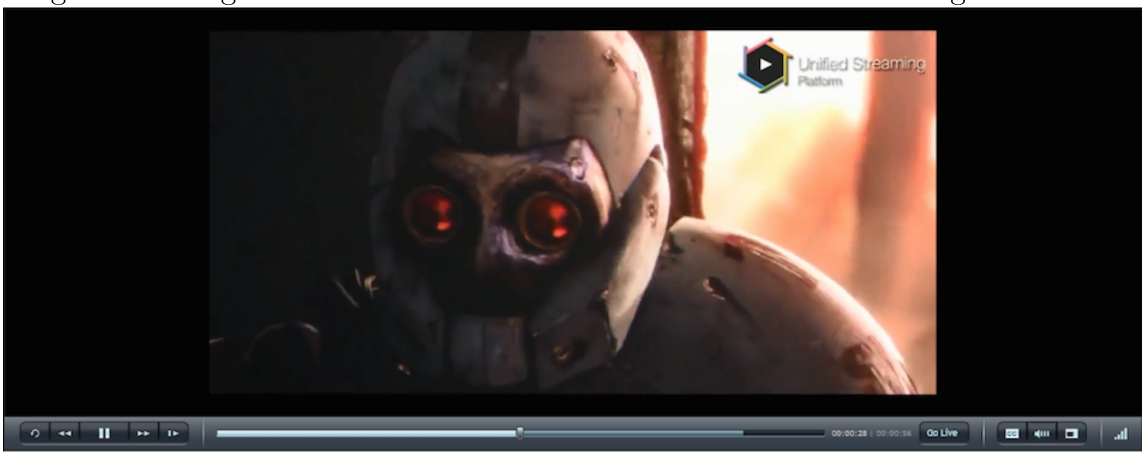

Figure 6.15: Stream Channel screenshot

A resume of the manifest of this video is shown in the next frame:

QualityLevel Index="0" Bitrate="412000" MaxWidth="400" MaxHeight="224"
QualityLevel Index="1" Bitrate="823000" MaxWidth="704" MaxHeight="396"
QualityLevel Index="2" Bitrate="2070000" MaxWidth="1280" MaxHeight="720"

It is possible to see the bitrates (the unit is bit/s) available for the video and the Width and Height for those bitrates. With the variation of the bandwidth available to Silverlight and probe, it is expected that both adapt to the network conditions and choose the adequate bitrate to stream. In this specific case, the scenario with downlink bandwidth 512 kbit/s is not shown, since the minimum bitrate is 412 kbit/s and it can not be displayed with this bandwidth (the video is stopped to buffer).

6.2.3.1 Streaming channel - Downlink 1024 kbit/s and Uplink 256 kbit/s

For this test scenario (similarly to the tests done before), the conditions of the network were an uplink bandwidth of 256 kbit/s and a downlink bandwidth of 1024 kbit/s.

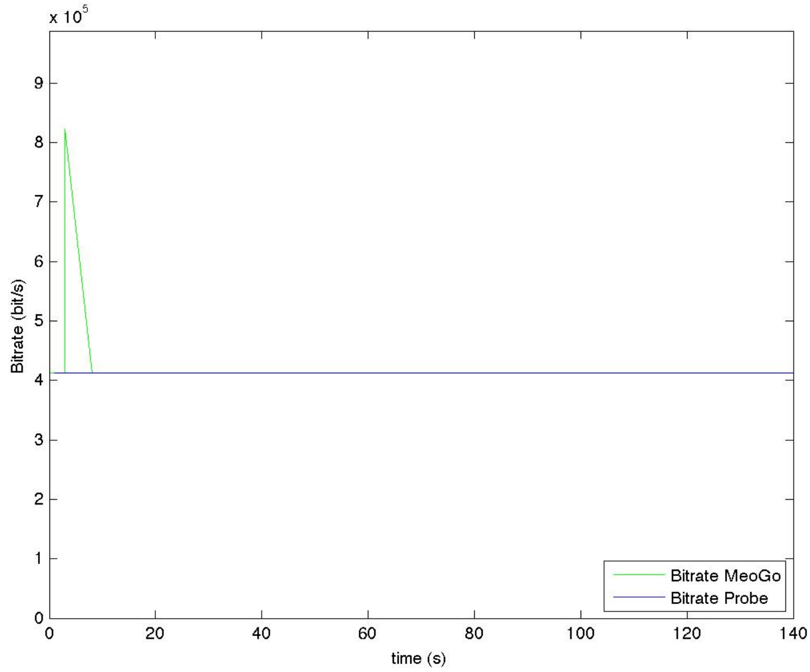

Figure 6.16: Bitrate for Streaming channel with 1024 kbit/s for downlink

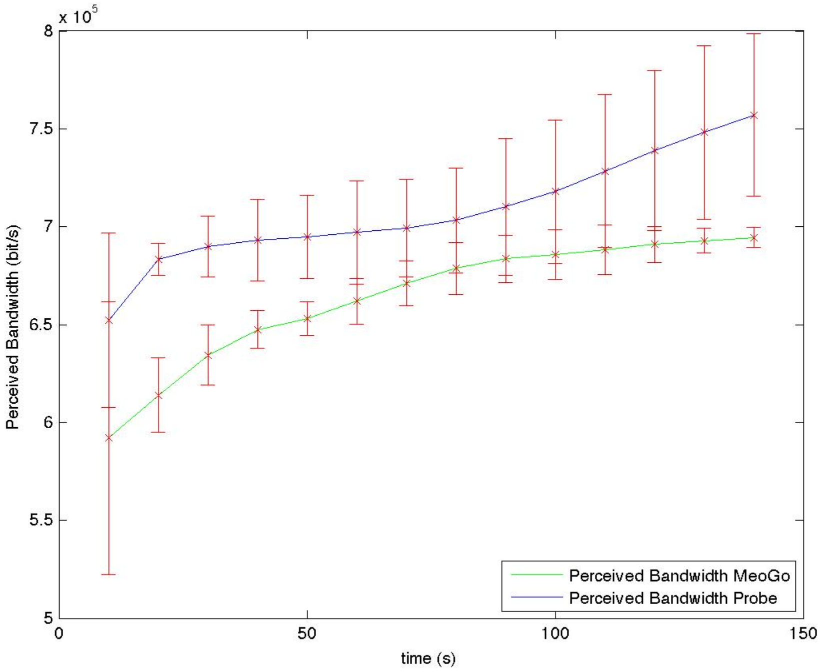

Figure 6.17: Perceived Bandwidth for Streaming channel with 1024 kbit/s for downlink

In figure 6.16, it is possible to see the chosen bitrate by boht probe and Silverlight.

Both processes quickly choose the lowest bitrate available 412 kbit/s. The peak in the Meo Go bitrate is a try to upgrade the bitrate, but it immediately finds that it is a mistake an returns to the lower bitrate.

The perceived bandwidth is shown in figure 6.17 and shows that the perceived bandwidth stays almost all the time between 600 kbit/s and 750 kbit/s, justifying the adoption of the 412 kbit/s bitrate (the next level for bitrate is 823 kbit/s that is higher than the perceived bandwidth). Relatively to the probe and Silverlight evaluation of bandwidth, the results are very similar becoming even closer with time.

6.2.3.2 Streaming channel - Downlink 10000 kbit/s and Uplink 5000 kbit/s

The network conditions to this scenario were an uplink bandwidth of 5000 kbit/s and a downlink bandwidth of 10000 kbit/s. It was requested to achieve the quality of service available in the network.

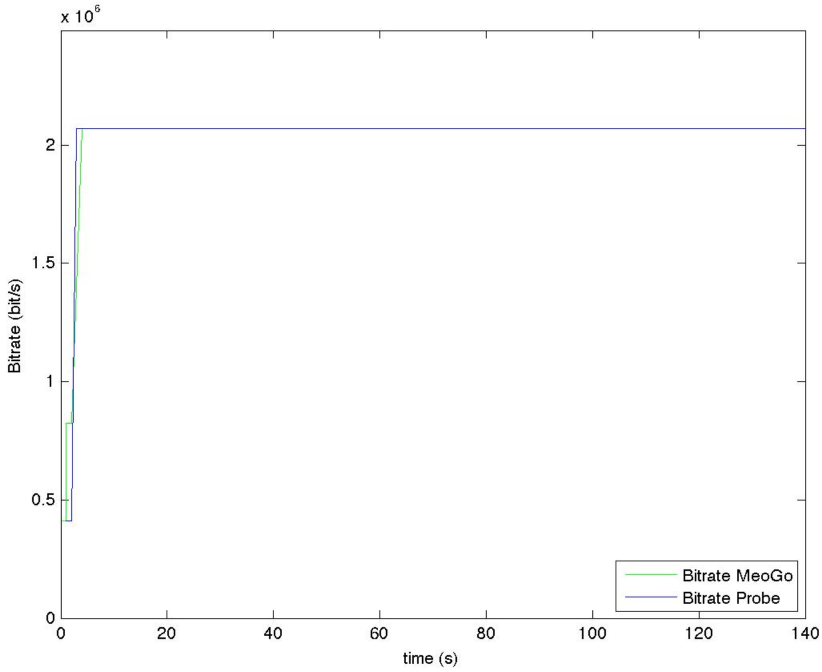

Figure 6.18: Bitrate for Streaming channel with 10000 kbit/s for downlink

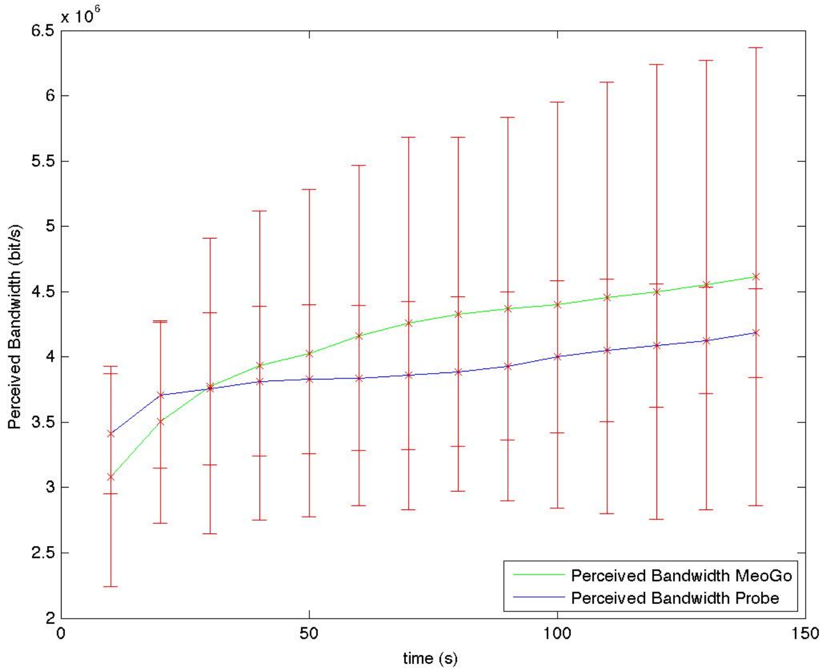

Figure 6.19: Perceived Bandwidth for Streaming channel with 10000 kbit/s for downlink

The chosen bitrate for the stream is shown in figure 6.18. In this case, both modules assume the higher bitrate available in a short period of time (the 2.07 Mbit/s level). This increases the results of the values of bandwidth perceived by both modules. Shown in figure 6.19, the perceived bandwidth has values between 6 Mbit/s and 7.5 Mbit/s, which allows the bitrate to be the maximum available (since it only requires 2.07 Mbit/s it can clearly choose the higher bitrate). The results from probe and Silverlight are close between each other; the confidence intervals intersect each other in some points and they are close in the remaining ones.

6.2.4 Comparison of Perceived Bandwidth for different Videos

In this section, it is made a comparison between the evaluation of the perceived bandwidth made by the three videos shown before (Mix video, Big bunny video and stream channel).

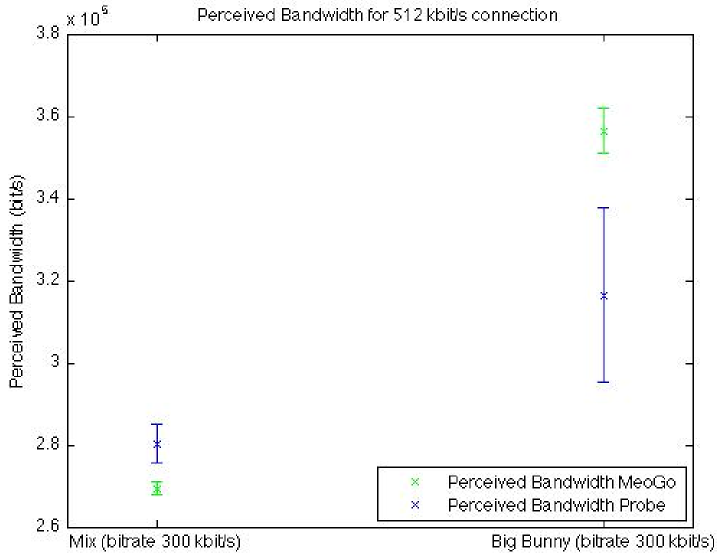

Figure 6.20: Perceived Bandwidth for 512 kbit/s

In figure 6.20, it is shown the evaluation of the Meo Go probe and the Meo Go service when the downlink bandwidth is 512 kbit/s. There are considered two videos, Mix video and Big bunny video (the stream video can not be displayed with a 512 kbit/s connection as was explained before). The results show that the probe and Meo Go make different evaluations of the perceived bandwidth when playing different videos. However, the values are not far from each other, having a maximum interval of 100 kbit/s. This can be justified by the use of different servers that are transmitting the different videos.

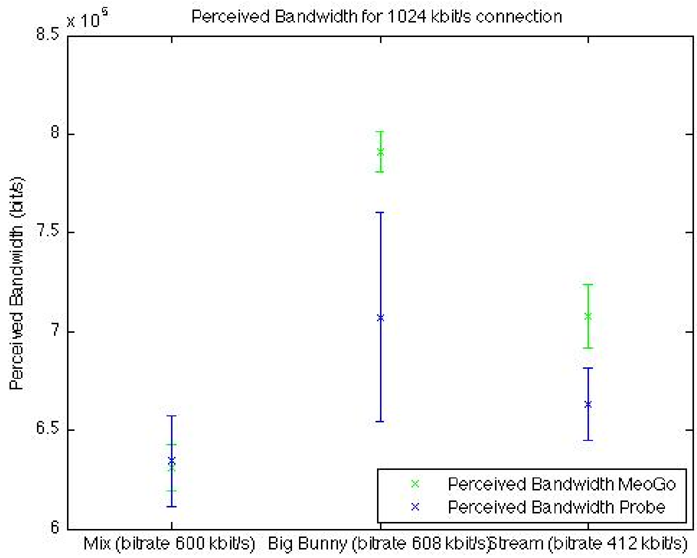

Figure 6.21: Perceived Bandwidth for 1024 kbit/s

In figure 6.21, it is shown the evaluation of the Meo Go probe and the Meo Go service when the downlink bandwidth is 1024 kbit/s. There are considered three videos: Mix video and Big bunny video and finally the stream video. It is possible to see that the results are close to each other being in a range of 200 kbit/s. Once again, the differences can be justified by the servers that provide the videos (that are different to each other), and the different bitrates used on the different videos (that are adapted to the network conditions), that cause the chunks to have different size in bits and therefore a different processing.

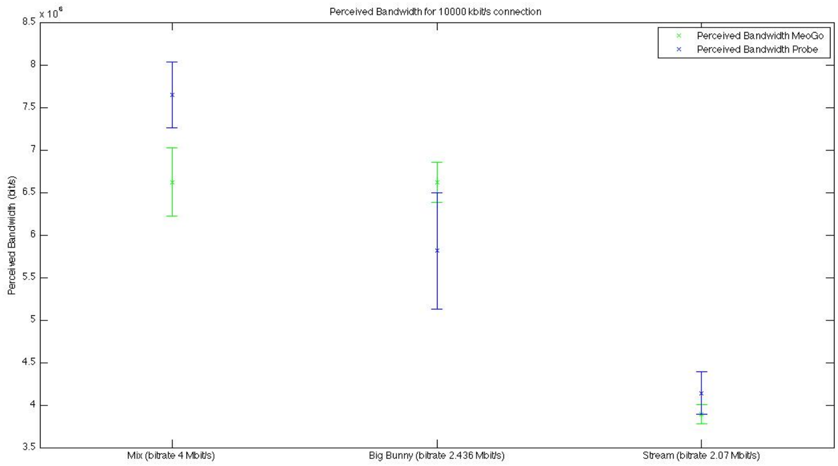

Figure 6.22: Perceived Bandwidth for 10 Mbit/s

The results for the perceived bandwidth with a 10 Mbit/s downlink connection comparing different videos is shown in figure 6.22. The videos used are the same as in the configuration with 1024 kbit/s shown before. It is possible to realize that the probe and Meo Go have different assessments of the perceived bandwidth when videos vary. The different videos with their different bitrates seem to cause, both in Meo Go and probe, to access differently the perceived bandwidth. This can be caused by the videos (that come from different servers) and also the different bitrates that cause both modules to process different amounts of data. However, both perceived bandwidth with MeoGo and the probe provide very similar results.

6.3 Testing Scenarios with Seagull clients

In this section it will be shown the results of the module that determines the congestion on the network and the PCRF resolution for congestion. This module is in charge of receiving the RAN metrics and be able to evaluate them into one qualitative value. After the report has been received by PCRF, and according with its value, it will impose in the network different policies.

The metrics used in these scenarios are extracted from one day of PT's RAN functioning, therefore they are real cases of congestion in the network. For these scenarios, three types of users will be considered in the cell. These users are divided in three categories:

- Gold client - The client with more privileges (higher bandwidth);
- Silver client - The client with medium privileges (medium bandwidth);
- Bronze client - The client with lower privileges (lower bandwidth);

In IpRaft solution, the downgrade of the clients is done by allocating new categories to the clients; for example, if a downgrade is supposed to be made on a Gold client, the possibilities are: allocate a silver plan, allocate a bronze plan or drop the client of the cell. The effects of the actions taken on the clients are possible to see in the package that is installed to the client (that is changed when it is chosen to be downgraded). The criteria to choose the clients that are going to be downgraded is shown in section 4.4, that says that the clients with better package must be protected; therefore, the first clients to be downgraded are the bronze, after them the silver and finally the gold.

The level of congestion determined by the Metrics Evaluation module is explained in the algorithm presented in section 4.2.
The number of sessions to downgrade is chosen based on the level of congestion given by the Metrics Evaluation module. The ruleset that determines the number of subscribers to downgrade was described in section 4.4, although it has now some adaptations. The ruleset implemented that the number of sessions to downgrade in a cell were 5%, 10% and 20% for the received levels of congestion 1, 2 and 3 respectively. Since there are only three users in the cell, the effects of different levels of congestion will always be the same (downgrade only one user, the rounding is made using rounding up to the next integer). Therefore, the levels of congestion have now new weights on the number of sessions to downgrade. The congestion level 1 corresponds to 5% of users in the cell (downgrading one client); the congestion level 2 corresponds to a downgrade of 50% of the total sessions in the cell (in practice, two sessions, since it is rounded up to the next integer); and finally, the congestion level 3 corresponds to 90% of the sessions in the cell (corresponding to three users to be downgraded in this case).

6.3.1 Scenario 1: Code Congestion

The content of the table is received by Metrics Evaluation module to acquire the level of congestion of the cell.

Data	Vendor	NodeB	Code Tree Usage Mean (%)	Code Fail (%)
2014-04-07 00:00	ERICSSON	U51004	88,52	0,43

Table 6.1: Scenario 1: Level of congestion

The metrics evaluation module calculates the level of congestion to send to PCRF.
Congestion value : Medium Congestion - value 2;
Cell ID : U51004;

On the side of PCRF, there are connected three seagull clients to IPM, that have all sent the initialize request that was guaranteed. They have the following packages, shown in table 6.2.

Seagull client	Package
961000001	Gold
961000002	Silver
961000003	Bronze

Table 6.2: Scenario 1: list of clients in the cell

The results of Metrics Evaluation Module that are received in the PCRF and its actions are the following:

Str [Congestion] = [2.0]
2014-06-30 17:42:35,791 Trace INFO — calling ruleSetName: congestionResolution
2014-06-30 17:42:35,800 IPMDefaultSS DEBUG — Number of sessions to modify:2

It is possible to see that PCRF received the congestion level 2, and determines that 2 sessions must be downgraded (as was explained in the introduction of this section). Then, PCRF sends an event informing that user 961000002 and 961000003 must be downgraded. Then, it is set the CCA RAR CONFIRMATION, confirming that the users' packages have been changed. The actual configuration in the cell is shown in table 6.3.

Seagull client	Package
961000001	Gold
961000002	Bronze

Table 6.3: Scenario 1: list of clients in the cell after PCRF policies

The client 961000003 is excluded from the cell (it was the expected, since the criteria used by PCRF determines that the bronze clients must be the ones to be downgraded first). The client 961000002 is downgraded (from Silver to Bronze), which is correct taking into account the criteria used.

6.3.2 Scenario 2: CE Congestion

The content of the table is received by Metrics Evaluation module to acquire the level of congestion of the cell.

Data	Vendor	NodeB	CE Cong Setup UL (%)	CE Cong Setup DL (%)	CE Usage Mean DL (%)	CE Usage Mean UL (%)
2014-04-07 00:00	ERICSSON	U51004	64,38	64,38	13,71	99,66

Table 6.4: Scenario 2: Level of congestion

The metrics evaluation module calculates the level of congestion to send to PCRF.
Congestion value : High Congestion - value 3;
Cell ID : U51004;

On the side of PCRF, there are connected three seagull clients to IPM, that have all send the initialize request and was guaranteed. They have the following packages, present in table 6.5.

Seagull client	Package
961000001	Gold
961000002	Silver
961000003	Bronze

Table 6.5: Scenario 2: list of clients in the cell

The results of Metrics Evaluation Module that are received in PCRF are:

Str [Congestion] = [3.0]
2014-06-30 17:45:26,167 Trace INFO — calling ruleSetName: congestionResolution
2014-06-30 17:45:26,177 IPMDefaultSS DEBUG — Number of sessions to modify:3

After this information, PCRF sends an event informing that all users must be downgraded.
Then the CCA RAR CONFIRMATION is sent, confirming that the users' packages have been changed. The actual configuration in the cell is shown in table 6.6.

Seagull client	Package
961000001	Silver
961000002	Bronze

Table 6.6: Scenario 2: list of clients in the cell after PCRF policies

The client 961000003 was excluded from the cell as was expected; the client 961000001 was downgraded, from Gold to the Silver package and the Silver client was downgraded from Silver to Bronze. This happened according to the policies configured, that determine that the level 3 of congestion must be done the downgrade of 3 sessions.

6.3.3 Scenario 3: Call Congestion PS

The content of the table is received by Metrics Evaluation module to acquire the level of congestion of the cell.

Data	Vendor	NodeB	Call CongestionPS (%)
2014-04-07 00:00	ERICSSON	U51004	50,00

Table 6.7: Scenario 3: level of congestion

The metrics evaluation module calculates the level of congestion to send to PCRF.
Congestion value : Medium Congestion - value 3;
Cell ID : U51004;

On the side of PCRF, are connected three seagull clients to IPM, that have all send the initialize request and was guaranteed. They have the following packages, shown in table 6.8.

Seagull client	Package
961000001	Gold
961000002	Silver
961000003	Bronze

Table 6.8: Scenario 3: list of clients in the cell

The results of Metrics Evaluation Module that are received in PCRF are:

```
Str [Congestion] = [3.0]
2014-06-30 17:47:38,867 Trace INFO — calling ruleSetName: congestionResolution
2014-06-30 17:47:38,874 IPMDefaultSS DEBUG — Number of sessions to modify:3
```

After this information, PCRF sends an event informing that all the users must be downgraded.
After the CCA RAR CONFIRMATION, confirming that the users' packages have been changed. The actual configuration in the cell is shown in table 6.9.

Seagull client	Package
961000001	Silver
961000002	Bronze

Table 6.9: Scenario 3: list of clients in the cell after PCRF policies

The client 961000003 was excluded from the cell and the two remaining clients were downgraded. This happened according the policies imposed, since congestion level 3 implies the downgrade of three users.

6.4 Testing Scenarios with Simulated Clients

In this section, it is shown the operation of the work developed in the RAN congestion detection and its resolution. As described before, the solution evaluates metrics from the RAN and gives a value of the congestion experienced on a cell to CAC/PCRF, that will act on the users in the cell and reduce the cell usage. In figure 6.23, it can be seen the architecture of the solution implemented.

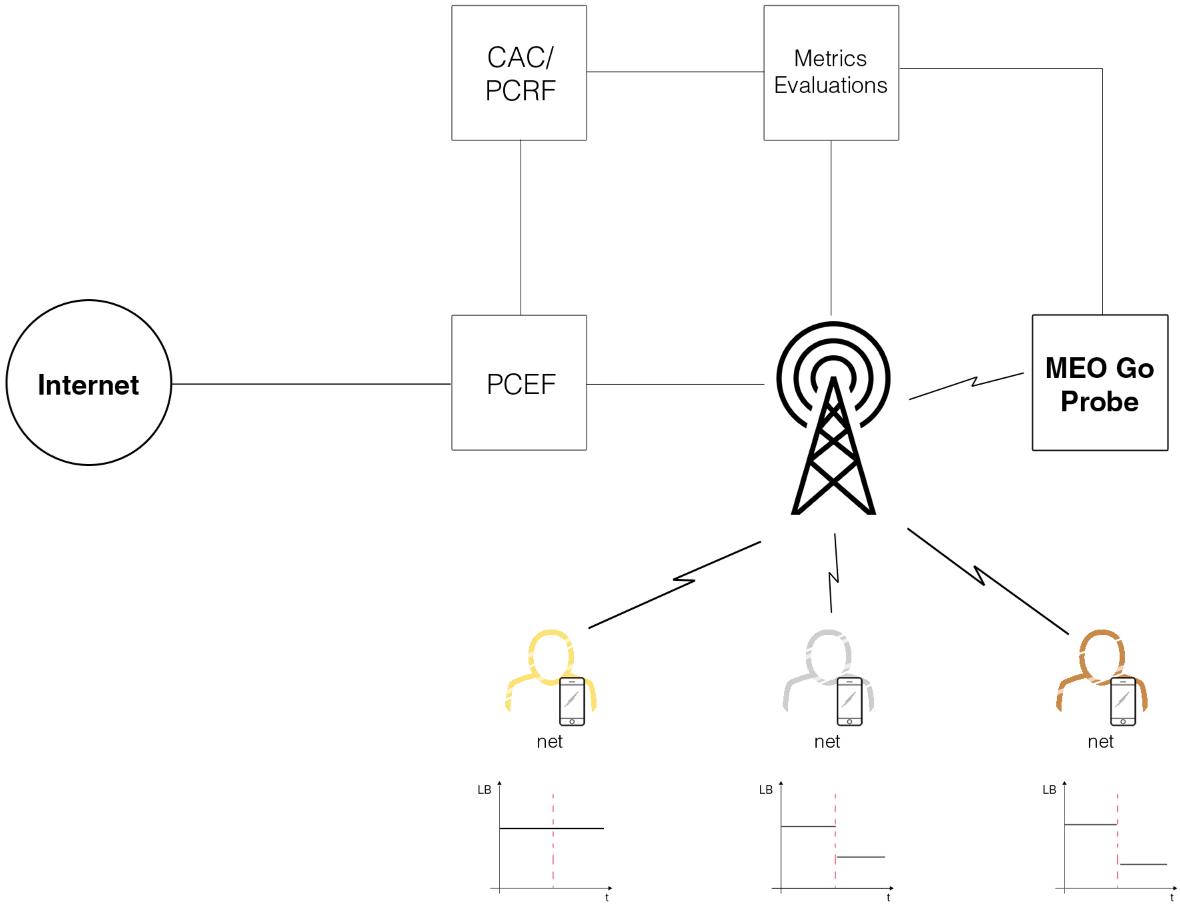

Figure 6.23: Solution Architecture

The metrics achieved by the Metrics Evaluation module are from Altaia from a day of usage of the PT's RAN network (they are not real time metrics achieved during the tests, since it was not possible to have all the equipment). The metrics from that specific day are evaluated, and the users in these cases will be affected like if they were in the cell evaluated. To perform the tests, three virtual machines (VMs from VirtualBox [46]) are used, that have the exact behaviour of an user with UE connected to the network. In this case, they perform the download of a file, "linux.iso", to generate traffic. The PCEF described in the figure 6.23 is a real network equipment from PT Inovação laboratory, named Cisco SCE 8000 [20]. This is a solution from Cisco based on deep packet inspection that is capable of detecting and classifying sessions, and also control IP traffic per subscriber. To evaluate the usage of bandwidth by the clients, it is used the NetPerSec application that is capable of measuring the traffic done by TCP/IP and giving information on the uplink and downlink instantaneous and average bandwidth usage. Finally, there are three packages available for the users: Gold, Silver and Bronze. The users can only change between these three packages or be dropped (being without any bandwidth connection). The description of the packages are:

- Gold client - The client with more privileges (higher bandwidth in figure 6.24), Max Downlink Bandwidth = 20 Mbit/s - Max Uplink Bandwidth = 20 Mbit/s;

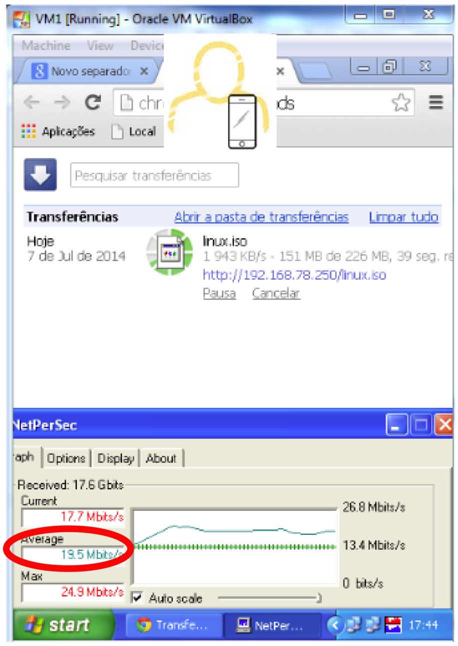

Figure 6.24: Downlink traffic generated by a Gold user (near 20 Mbit/s)

- Silver client - The client with medium privileges (medium bandwidth in figure 6.25), Max Downlink Bandwidth = 4 Mbit/s - Max Uplink Bandwidth = 4 Mbit/s;

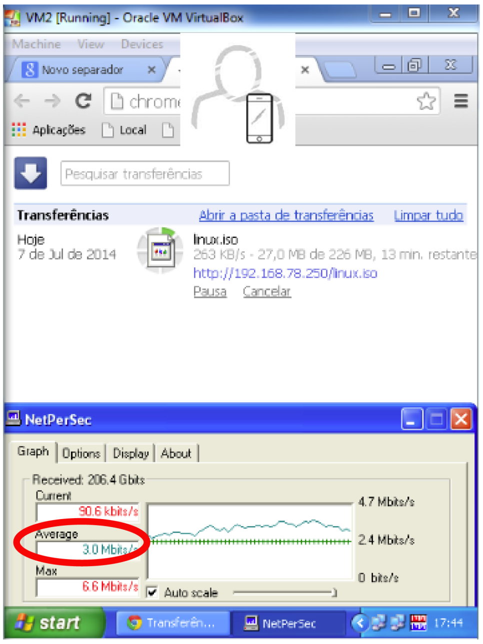

Figure 6.25: Downlink traffic generated by a Silver user (near 4 Mbit/s)

- Bronze client - The client with lower privileges (lower bandwidth in figure 6.26), Max Downlink Bandwidth = 1 Mbit/s - Max Uplink Bandwidth = 1 Mbit/s;

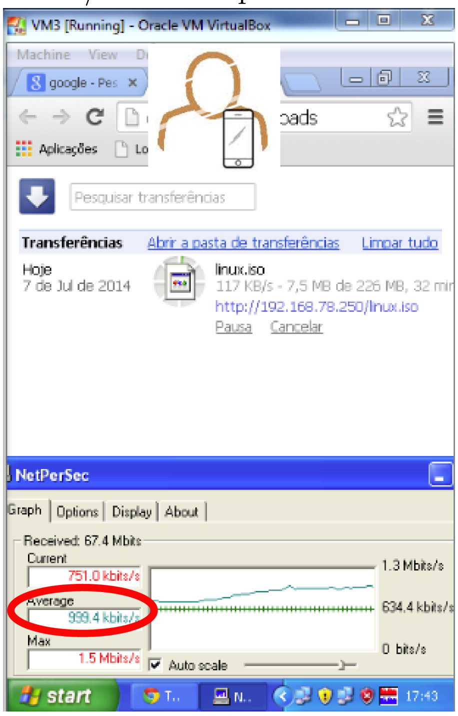

Figure 6.26: Downlink traffic generated by a Bronze user (near 1 Mbit/s)

The criteria to determine which percentage of users in the cell will be downgraded is the same as in the tests made with the seagull test (explained in section 6.3). Summarizing, the congestion level 1 corresponds to 1 user downgraded, congestion level 2 corresponds to 2 users downgraded and congestion level 3 corresponds to 3 users downgraded. The order to choose which client is going to be downgraded first was also explained in section 6.3 and 4.2.

6.4.1 Scenario 1: Code Congestion

In a cell there are three users: one with the Gold package, one with the Silver package and the last one with the Bronze package. It is inspected a report from Altaia containing information on the behaviour of a cell. The actuation of Metrics Evaluation Module can be seen in figure 6.27.

```
START
______________________________________________________________________
| ###### Report ######
|There is congestion
|NodeID -> U51004
----------------------------------------------------------------------
|Code Tree Usage Mean (%)
|Parameter Value: 88.524167 higher than the minimum threshold: 80.0
|Max Congestion value registed for the cell: 2.0
----------------------------------------------------------------------
```

Figure 6.27: Metrics Evaluation Module detecting code congestion

With the evaluation of the cell, it is possible to see that there is congestion on the cell and it is due to the code resource occupation. The value of congestion is 2 (due to the process of Metrics Evaluation module explained in Section 4.2), and it is then sent to CAC/PCRF so it can apply the correct rules to the users. After running the ruleset that deals with the RAN congestion, it is determined that two users must be downgraded (the process is explained in Section 4.4). Since the objective is to "protect" the users with higher package, the ones to be downgraded are the ones with lower package. Therefore, the Silver and Bronze users are going to be downgraded.

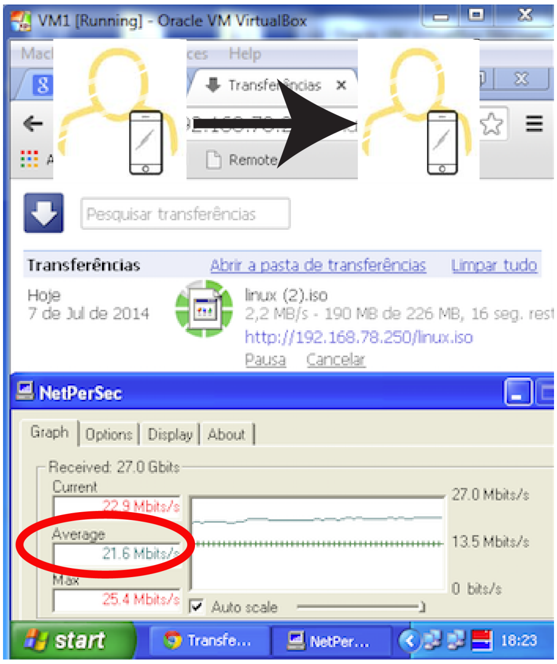

Figure 6.28: Scenario 1 user 1

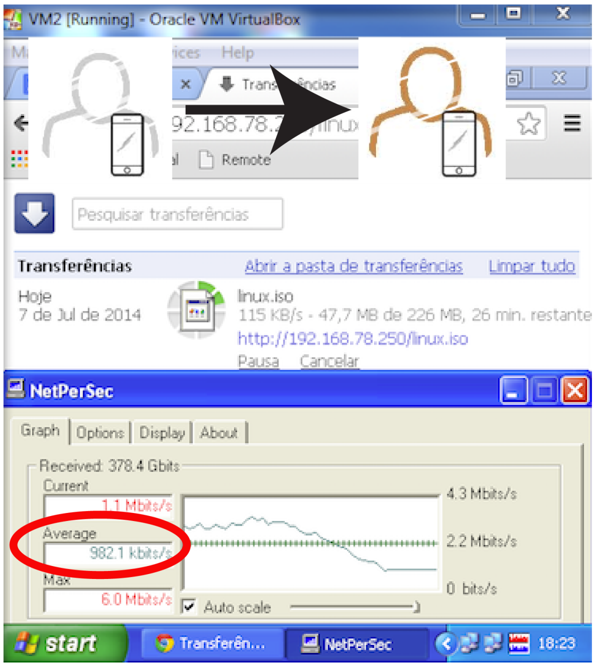

Figure 6.29: Scenario 1 user 2

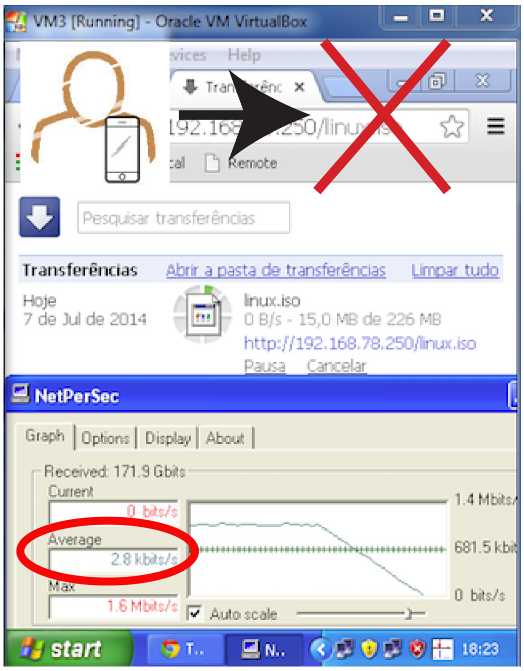

Figure 6.30: Scenario 1 user 3

In figure 6.27, it is possible to see that the Metrics Evaluation module does the right evaluation of the cell (according with process defined in Section 4.2) determining the value 2 for congestion. In figures 6.29 and 6.30, it is possible to observe that the Silver and

Bronze users are downgraded, the Silver user was downgraded to the Bronze package (with 1 Mbit/s), while the initial Bronze user was dropped from the cell (the bandwidth is close to 0 Mbit/s). The Gold user maintains is package and bandwidth without any modification, that can be seen in figure 6.28.

6.4.2 Scenario 2: Channel Elements Congestion

The users in the cell are the same that in the previous scenario: one Gold, one Silver and one Bronze. It is inspected a report from Altaia containing information on the behaviour of a cell. The actuation of Metrics Evaluation Module can be seen in figure 6.31.

```
START
________________________________________________________________________
| ###### Report ######
|There is congestion
|NodeID -> U51004
------------------------------------------------------------------------
|CE Cong Setup UL (%)
|Parameter Value: 64.378967 higher than the minimum threshold: 5.0
|Max Congestion value registed for the cell: 3.0
------------------------------------------------------------------------
|CE Usage Mean UL (%)
|Parameter Value: 99.662001 higher than the minimum threshold: 80.0
|Max Congestion value registed for the cell: 3.0
------------------------------------------------------------------------
|CE Cong Setup DL (%)
|Parameter Value: 64.378967 higher than the minimum threshold: 5.0
|Max Congestion value registed for the cell: 3.0
------------------------------------------------------------------------
```

Figure 6.31: Metrics Evaluation Module detecting Channel Element congestion

With the evaluation of the cell done, it is possible to observe that there is congestion on the cell, and it is due to the channel elements occupation. The value of congestion obtained is 3 and it is then sent to CAC/PCRF, so it can apply the rules to the users. After running the ruleset that deals with the RAN congestion, it is determined that three users must be downgraded. Therefore, all the users in the cell must be downgraded.

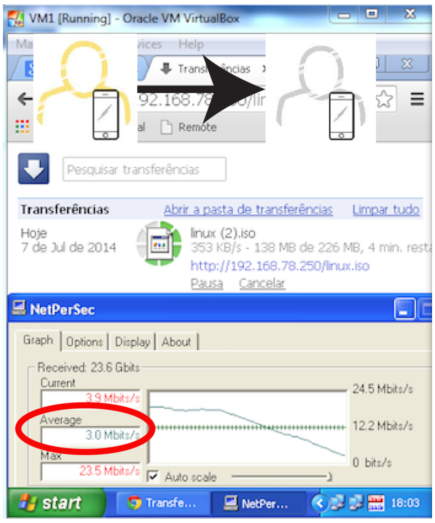

Figure 6.32: Scenario 2 user 1

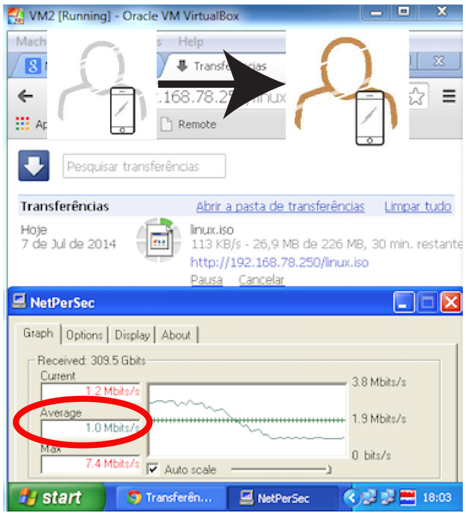

Figure 6.33: Scenario 2 user 2

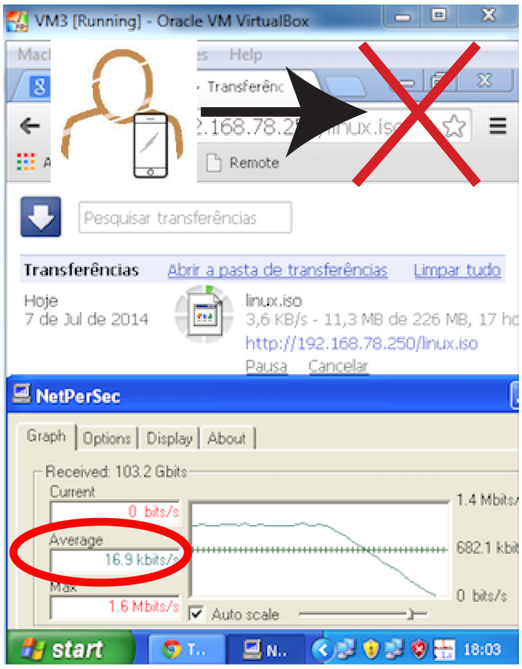

Figure 6.34: Scenario 2 user 3

In figure 6.31 is possible to observe that the Metrics Evaluation module does a correct evaluation of the parameters, achieving the right level of congestion for the cell (congestion level 3). In figures 6.32, 6.33 and 6.34 it is observed that every user is downgraded to a lower package compared to the one that they had before.

6.4.3 Scenario 3: Load Congestion

In a cell there are three users, two with the Silver package and one with the Bronze package. A report from Altaia containing information on the behaviour of a cell is inspected by Metrics Evaluation Module and the result can be seen in figure 6.35.

```
START
_______________________________________________________________________
| ###### Report ######
|There is congestion
|NodeID -> U51225
-----------------------------------------------------------------------
|Load Setup Fail (%)
|Parameter Value: 5.2 higher than the minimum threshold: 5.0
|Max Congestion value registed for the cell: 1.0
-----------------------------------------------------------------------
```

Figure 6.35: Metrics Evaluation Module detecting Load congestion

With the evaluation of the cell done, is possible to see that there is congestion on the cell, and it is due to the excessive load on the cell. The value of congestion is 1 and it is then sent to CAC/PCRF, so it can apply the rules to the users. After running the ruleset that deals with the RAN congestion, it is determined that one user must be downgraded.

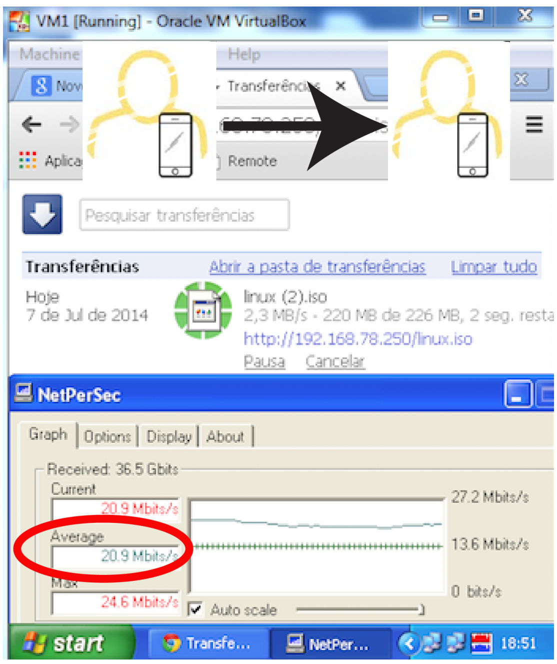

Figure 6.36: Scenario 3 user 1

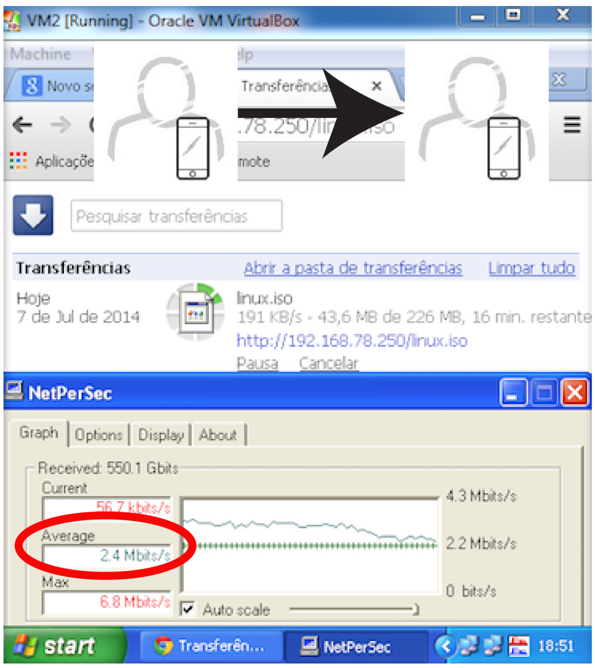

Figure 6.37: Scenario 3 user 2

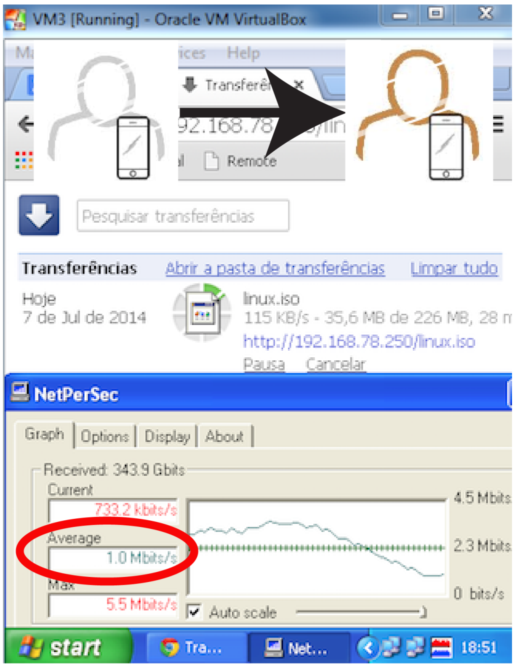

Figure 6.38: Scenario 3 user 3

In the figure 6.35, it is possible to observe that the evaluation module allocates the level of congestion equal to 1, which is the right procedure described in Section 4.2. Only one user is downgraded (in fact it is dropped due to the fact that is is a Bronze user, and the lowest level does not exist, figure 6.38), and the other two users remain with no change, as it is shown in figures 6.36 and 6.37.

6.4.4 Scenario 4: Meo Go problems

It is produced a report, containing a value for Meo Go service capable of being evaluated by Metrics Evaluation module. This report is done as follows: the probe has the same bandwidth of a Silver client on the cell (available bandwidth of 4 Mbit/s) and it is running the big bunny video.
The following are the defined values of QoS for Meo Go:

- Level 1 - medium quality - third bitrate available (608 kbit/s);
- Level 2 - medium/low quality - second bitrate available (427 kbit/s);
- Level 3 - low quality - lowest bitrate available (300 kbit/s).

Considering the bitrate achieved by the probe in this cell has the value of 608 kbit/s, the level attributed to the QoS is 1.

In this cell there are three users, two with the Silver package and one with the Bronze package. It is inspected a report containing information on the behaviour of the Meo Go.

The actuation of Metrics Evaluation Module can be seen in figure 6.39.

```
START
________________________________________________________________________
| ###### Report ######
|There is congestion
|NodeID -> U51550
------------------------------------------------------------------------
|Meo Go
|Parameter Value: 1.0 higher than the minimum threshold: 0.0
|Max Congestion value registed for the cell: 1.0
------------------------------------------------------------------------
```

Figure 6.39: Metrics Evaluation Module detecting problems in Meo Go service

With the evaluation of the cell done, is possible to see that there is problems in the Meo Go, causing the Metrics Evaluation Module to attribute the cell the level of congestion 1 (in this cases the level allocated to QoS is equal to the level of congestion). The value of congestion is then sent to CAC/PCRF, so it can apply the rules to the users. After running the ruleset that deals with the RAN congestion, it is determined that one users must be downgraded. Since the objective is to "protect" the user with higher package, the ones to be downgraded are the ones with lower package. Therefore, the Bronze user is going to be downgraded.

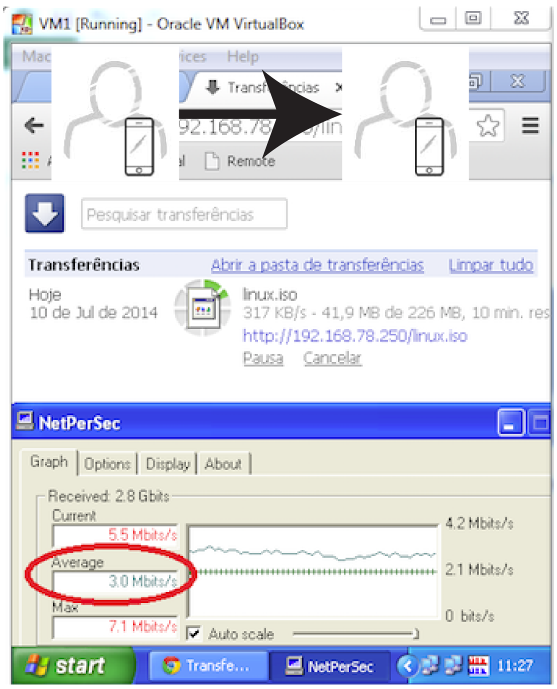

Figure 6.40: Meo Go User 1

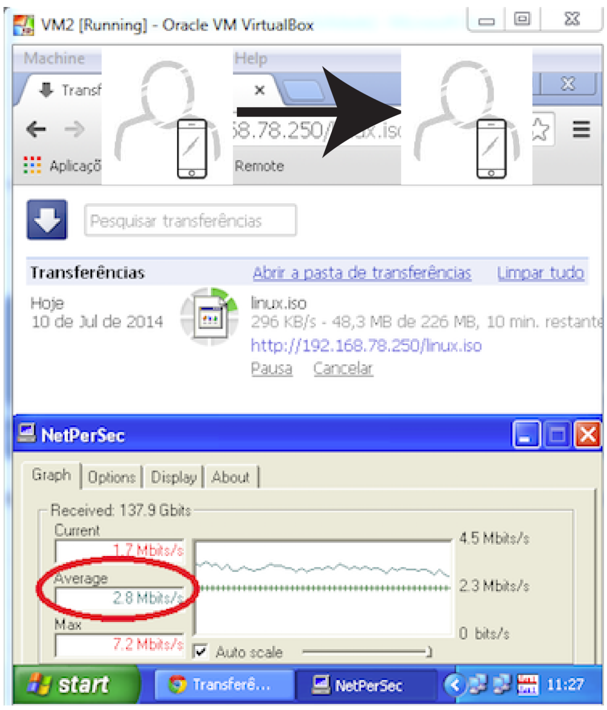

Figure 6.41: Meo Go User 2

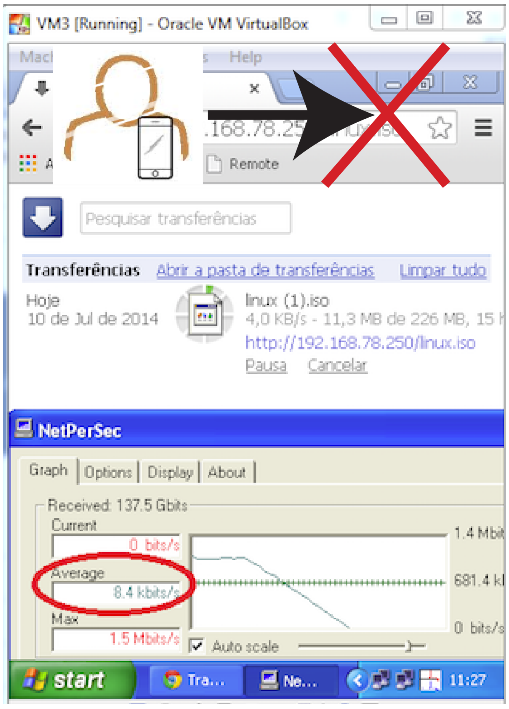

Figure 6.42: Meo Go user 3

In the figure 6.42, we observe that the Bronze user is dropped of the cell (its bandwidth is almost 0 kbit/s), and the other users in figures 6.41 and 6.40 maintain their original packages and consequently the same bandwidth.

6.5 Conclusions

This chapter has presented the results of the tests conducted on the different modules of the entire solution.
The first results shown the Meo Go probe service evaluation. The probe was tested using different scenarios that comprise variations in the bandwidth available and variations on the type of video streamed. The results are similar to the ones given by the original Meo Go service which show that the probe is able to emulate the MeoGo service with a good accuracy.
Then, it is shown the operation of the metrics evaluation module, working together with PCRF with the users simulated with Seagull. The behaviour of both modules were the expected and they perform their actions perfectly. The actuation of the policies in the Seagull users is done correctly, verified by the installation of the packages.
The last test scenarios, evaluated Metrics Evaluation module, PCRF and Meo Go Probe at the same time, with users simulated with virtual machines (simulating the behaviour of real clients on a service provider network). Different levels of congestion were tested with different types of users in the cell, to cover a large set of scenarios possible. It was possible

to see the actuation of the PCRF in users by observing the pictures of the bandwidth available to each user at the moment of actuation. The behaviour of all the modules was as expected, and they did their expected operation flawlessly.

Chapter 7

Conclusions and Future Work

7.1 Conclusions

The work developed in this Dissertation aimed to solve a known problem of radio access networks: congestion. For this purpose, several solutions were taken into account, and an architecture that considers both metrics from network management and monitoring systems and probing systens (both passive and active metrics) was specified, developed and tested. Probes can give specific information on users/services experience on a cell while passive metrics can give the state of the cell. The chosen way to detect congestion was the evaluation of radio access network parameters, that were obtained via an PT Inovação platform Altaia, while the perception of the QoS experienced in the network for specific services was provided using probes.

After understanding the existing architecture of PT Inovação IpRaft, it was specified an architecture to the solution, which contains a new module, the metrics evaluation module, that is capable to connect to PCRF and warn it that congestion is occurring. Simultaneously, it was developed a probe for the Meo Go service that can give the QoS of the services on a specific cell. With the information provided by the modules described, PCRF is now able to choose and create dynamic rules to resolve cases of congestion in the network.

The defined use cases were tested using the integration of the modules with the existing architecture. The Meo Go was tested in different types of scenarios, with different videos and bandwidth conditions. The metrics evaluation module and PCRF were tested, first using Seagull emulated clients, where it is possible to see the detection of congestion for reports taken from Altaia and observe the actions of the PCRF to the users in the case of congestion. Then, tests were performed using clients simulated with virtual machines, that have a behaviour very similar to a real client, being able to generate traffic. It was possible to test the metrics evaluation module working combined with the probes and PCRF, and finally observe the actions taken on the users with real reductions of their bandwidth (proved with the observed reduction of download throughput).

From the obtained results, it is possible to conclude that the various modules of the solution are capable of solving the congestion problems with a good performance. The metrics evaluation module is able to detect congestion based on the evaluation of specific parameters and combine them to achieve a general qualitative value to congestion. In the probe results, it is possible to observe that the probe has a similar behaviour comparing to the original Meo Go architecture having very good results for bitrate and perceived bandwidth. In the case of CAC decision, we showed that it is able to implement the rules defined, with different actions on the users in the cell, according with the received quality and congestion levels, and according to the users and services profiles.

This architecture and modules is integrated in the PT Inovação test network, and will be able to be included in a future release of their products.

7.2 Future Work

During the work of this Dissertation, it was possible to observe some possible future improvements to the proposed solution:

- Use this implementation to other technologies: 3G is still commonly used, but 4G is starting to increase its usage that will lead to congestion. It would be important if the solution created in this Dissertation was expanded to the 4G technology congestion detection.

- The creation of probes for other types of interesting services (like Youtube) from the point of view of the operator would create a richer solution to the problem of congestion or even creating new forms of managing the network for these very common services.

- The integration of probes informing of events from the low level information of the network in the existing solution will also be advantageous. Information like incoming handovers to the congested cell, that can be achieved through probing will help the ipRaft solution to avoid and stop congestion intelligently.

- Force the handover of users to other technologies: one of the solutions studied to solve congestion is to force the handover of users to other technologies that are not congested. With the use of ANDSF it is possible to suggest the user equipment to change the radio access technology, and in fact avoid congestion and make use of existing technologies and infrastructures.

Bibliography

[1] 3GPP TS 23.002. Network architecture (release 12.4.0). http://www.3gpp.org/ftp/Specs/archive/23_series/23.002/23002-c40.zip, 2014. [accessed 12-June-2014].

[2] 3GPP TS 23.203. Policy and charging control architecture (release 12.4.0). http://www.3gpp.org/ftp/Specs/archive/23_series/23.203/23203-c40.zip, 2014. [accessed 09-June-2014].

[3] 3GPP TS 25.401. Utran overall description (release 12.0.0). http://www.3gpp.org/ftp/Specs/archive/25_series/25.401/25401-c00.zip, 2013. [accessed 15-June-2014].

[4] Jens Fiedler Thomas Magedanz Adel Al-Hezmi, Marius Corici. Network resources pre-reservation for multimedia content delivery in heterogeneous wireless networks. *Wireless Conference 2011 - Sustainable Wireless Technologies (European Wireless), 11th European*, 2011.

[5] Adobe. Http dynamic streaming. http://www.adobe.com/products/hds-dynamic-streaming.html, 2014. [accessed 10-June-2014].

[6] E.A. Aghdaee, Monash University. Department of Electrical, and Computer Systems Engineering. *Quality of Service Support in IEEE 802.16 Broadband Wireless Access Networks*. 2006.

[7] Y. Ahmet Sekercioglu, Milosh Ivanovich, and Alper Yegin. A survey of MAC based QoS implementations for WiMAX networks. *Computer Networks*, 53:2517–2536, 2009.

[8] Alcatel-Lucent. 5780 dynamic services controller. http://www.alcatel-lucent.com/products/5780-dynamic-services-controller-dsc, 2013. [accessed 17-June-2014].

[9] Alcatel-Lucent. 9900 wireless network guardian. http://www.alcatel-lucent.com/products/wireless-network-guardian, 2013. [accessed 17-June-2014].

[10] Alcatel-Lucent. Intelligent traffic management. http://www.alcatel-lucent.com/products/5780-dynamic-services-controller-dsc, 2013. [accessed 17-June-2014].

[11] Allot. Dart. http://www.allot.com/Dynamic_Actionable_Recognition_Technology.html, 2014. [accessed 10-June-2014].

[12] Amdoc. Amdocs policy controller. http://www.amdocs.com/products/network-control/3G-LTE-control-plane/Pages/amdocs-policy-controller.aspx, 2013. [accessed 4-June-2014].

[13] Elan Amir. The case for deep packet inspection. http://www.eetimes.com/document.asp?doc_id=1280856, 2007. [accessed 10-June-2014].

[14] Luis Alonso Angelos Antonopoulos, Elli Kartsakli and Christos Verikoukis. *Recent Advances in Wireless Communications and Networks*, chapter Dealing with VoIP Calls During "Busy Hour" in LTE. 2011.

[15] Arieso. ariesogeo platform. http://www.arieso.com/products/platform/, 2013. [accessed 2-June-2014].

[16] Aruna Balasubramanian, Ratul Mahajan, and Arun Venkataramani. Augmenting mobile 3G using WiFi. In *Proceedings of the 8th international conference on Mobile systems, applications, and services (MobiSys)*, page 209, 2010.

[17] P Calhoun, J. Loughney, E. Guttman, G. Zorn, and J. Arkko. Diameter Base Protocol. *IETF RFC 3588*, pages 1–148, 2003.

[18] Simao Cardeal, Filipe Neves, Salviano Soares, Filipe Tavares, and Pedro Assuncao. ArQoS: System to monitor QoS/QoE in VoIP. In *EUROCON 2011 - International Conference on Computer as a Tool - Joint with Conftele 2011*, 2011.

[19] S. M. Chadchan and C. B. Akki. Priority-scaled preemption of radio resources for 3gpp lte networks. *International Journal of Computer Theory and Engineering*, 3(6), December 2011.

[20] Cisco. Cisco sce 8000 series service control engine. http://www.cisco.com/c/en/us/products/service-exchange/sce-8000-series-service-control-engine/index.html, 2013. [accessed 23-June-2014].

[21] Marius Iulian Corici, Fabricio Carvalho De Gouveia, and Thomas Magedanz. A network controlled QoS model over the 3GPP system architecture evolution. In *The 2nd International Conference on Wireless Broadband and Ultra Wideband Communications, AusWireless 2007*, 2007.

[22] Microsoft Corporation. Live and on-demand video with silverlight and iis smooth streaming. www.microsoft.com/silverlight/smoothstreaming/, February 2010. [accessed 10-June-2014].

[23] Antonio De La Oliva, Carlos J. Bernardos, Maria Calderon, Telemaco Melia, and Juan Carlos Zuniga. IP flow mobility: Smart traffic offload for future wireless networks. *IEEE Communications Magazine*, 49:124–132, 2011.

[24] Hannes Ekstrom. QoS control in the 3GPP evolved packet system. *IEEE Communications Magazine*, 47:76–83, 2009.

[25] Ericsson. Ce utilization calculation for ericsson. `http://www.academia.edu/4556879/CE_Utilization_Calculation_for_ERICSSON`, 2009. [accessed 10-June-2014].

[26] J Vijay Franklin and K Paramasivam. Article: Hybrid adaptive call admission control mechanism for ensuring qos in 3gpp: Lte networks. *International Journal of Computer Applications*, 42(21):36–41, March 2012. Published by Foundation of Computer Science, New York, USA.

[27] PT Inovação. Arqos. `http://www.openidea.com/portfolio/arqos/`, 2013. [accessed 17-June-2014].

[28] B. Sas K. Spaey and C. Blondia. Self-optimising call admission control for lte downlink. *SOCRATES project*, February 2010.

[29] S.K. Kasera, R. Ramjee, S.R. Thuel, and Xin Wang Xin Wang. Congestion control policies for IP-based CDMA radio access networks. *IEEE Transactions on Mobile Computing*, 4, 2005.

[30] Xiangyi Kong and Jihong Zhao. *Cross-layer-based mapping mechanism in LTE Advanced*. IEEE, 2011.

[31] Sarah Koskie and Zoran Gajic. A Nash game algorithm for SIR-based power control in 3G wireless CDMA networks. *IEEE/ACM Transactions on Networking*, 13:1017–1026, 2005.

[32] Dirk Kutscher, Henrik Lundqvist, and Faisal Ghias Mir. Congestion exposure in mobile wireless communications. In *GLOBECOM - IEEE Global Telecommunications Conference*, 2010.

[33] Young Min Kwon, Jun Suk Kim, Jaheon Gu, and Min Young Chung. Andsf-based congestion control procedure in heterogeneous networks. *Information Networking (ICOIN), 2013 International Conference on*, pages 547–550, January 2013.

[34] J. Laiho-Steffens, A. Wacker, and K. Sipila. Verification of 3G radio network dimensioning rules with static network simulations. *VTC2000-Spring. 2000 IEEE 51st Vehicular Technology Conference Proceedings (Cat. No.00CH37026)*, 1, 2000.

[35] Jani Lakkakorpi, Ove Strandberg, and Jukka Salonen. Adaptive connection admission control for differentiated services access networks. *IEEE Journal on Selected Areas in Communications*, 23:1963–1971, 2005.

[36] Haipeng Lei, Mingliang Yu, Anjun Zhao, Yongyu Chang, and Dacheng Yang. Adaptive connection admission control algorithm for LTE systems. In *IEEE Vehicular Technology Conference*, pages 2336–2340, 2008.

[37] Carlos Marques, Susana Sargento, Carlos Parada, Pedro Neves, Carlos Rodrigues, Miguel Santos, Tiago Cardoso, Telma Mota, Francisco Fontes, and Filipe Rodrigues. Controlo de admissão e reserva de recursos em ambientes 4g epc/lte. Technical report, IT Aveiro, 2013.

[38] Meo. Meo go. `http://meogo.meo.pt/oquee/Pages/produtos.aspx`, 2014. [accessed 10-June-2014].

[39] Microsoft. Visual studio. `http://www.visualstudio.com/`, 2013. [accessed 21-March-2014].

[40] Usama Mir and Loutfi Nuaymi. Comparison of policy realization strategies for lte networks. *Vehicular Technology Conference (VTC Fall), 2012 IEEE.*

[41] Movik. Reach. `http://www.movik.com/movik-introduces-reach%E2%84%A2-architecture-foundation-intelligent-ran`, 2012. [accessed 6-June-2014].

[42] J. Mueller. Ue and network initiated qos reservation in ngn and beyond. *Network of the Future (NOF), 2011 International Conference on the*, 2011.

[43] Szilveszter Nadas, Sandor Racz, Zoltan Nagy, and Sandor Molnar. Providing congestion control in the Iub Transport Network for HSDPA. In *GLOBECOM - IEEE Global Telecommunications Conference*, pages 5293–5297, 2007.

[44] NetFort. Languardian. `http://www.netfort.com/languardian/architecture/deep-packet-inspection`, 2012. [accessed 12-June-2014].

[45] D. Niyato and E. Hossain. Call admission control for qos provisioning in 4g wireless networks: issues and approaches. *Network, IEEE*, 19, 2005.

[46] Oracle. Virtualbox. `https://www.virtualbox.org/`, 2013. [accessed 21-June-2014].

[47] SA Portugal Telecom Inovação. Altaia. `http://www.ptinovacao.com.br/download.do?image=/pdf/ALTAIA.pdf`. [accessed 11-June-2014].

[48] SA Portugal Telecom Inovação. ip-raft. `http://www.ptinovacao.com.br/download.do?image=/pdf/IP-Raft.pdf`. [accessed 14-June-2014].

[49] Procera. Procera navl. `http://www.proceranetworks.com/products/deep-packet-inspection-navl`, 2014. [accessed 1-June-2014].

[50] Qosmos. Qosmos deepflow: Quality of experience use case. http://www.qosmos.com/new-qosmos-deepflow-probes-deliver-the-power-of-network-intelligence-in-ready-to-2014. [accessed 2-June-2014].

[51] Qosmos. Qosmos deepflow: Subscriber analytics use case. http://www.qosmos.com/wp-content/uploads/2014/01/Qosmos-DeepFlow-Analytics-use-case-datasheet-Jan-2014.pdf, 2014. [accessed 10-June-2014].

[52] Angelos N. Rouskas and Dimitrios N. Skoutas. OVSF codes assignment and reassignment at the forward link of W-CDMA 3G systems. In *IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, PIMRC*, volume 5, pages 2404–2408, 2002.

[53] Sandvine. Policy engine. https://www.sandvine.com/platform/policy-engine.html, 2013. [accessed 2-June-2014].

[54] Nishit Narang Sumit Kasera. *3G Networks*. Tata McGraw-Hill Education, 2004.

[55] Tekelec. Improving the customer experience with ran-aware policy management. http://www.tekelec.com, 2011. [accessed 8-June-2014].

[56] Elias Z. Tragos, George Tsiropoulos, George T. Karetsos, and Sofoklis A. Kyriazakos. Admission control for QoS support in heterogeneous 4G wireless networks, 2008.

[57] B. Wehbi and E.M. Sankala, J. ; de Oca. Network monitoring challenges in the evolved packet core. *Future Network and Mobile Summit (FutureNetw)*, 2012.